

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
22 February 2001 (22.02.2001)

PCT

(10) International Publication Number  
**WO 01/12819 A2**

(51) International Patent Classification<sup>7</sup>: C12N 15/55,  
9/16, C07K 16/40, C12Q 1/42, 1/68, A61K 38/46

(21) International Application Number: PCT/US00/22158

(22) International Filing Date: 11 August 2000 (11.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/149,005 13 August 1999 (13.08.1999) US

(63) Related by continuation (CON) or continuation-in-part  
(CIP) to earlier application:  
US 60/149,005 (CIP)  
Filed on 13 August 1999 (13.08.1999)

(71) Applicant (for all designated States except US): SUGEN,  
INC. [US/US]; 230 East Grand Avenue, South San Francisco,  
CA 94080 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): PLOWMAN,  
Gregory, D. [US/US]; 4 Honeysuckle Lane, San Carlos,  
CA 94070 (US). MARTINEZ, Ricardo [US/US]; 984  
Cartier Lane, Foster City, CA 94404 (US). WHYTE,  
David [US/US]; 2623 Barclay Way, Belmont, CA 94002  
(US). HILL, Ron [US/US]; Sugent, Inc., 230 East Grand

Avenue, South San Francisco, CA 94080 (US). FLANAGAN,  
Peter [US/US]; 192 Liberty Street, San Francisco,  
CA 94110 (US). LIOUBIN, Mario [US/US]; Sugent, Inc.,  
230 East Grand Avenue, South San Francisco, CA 94080  
(US).

(74) Agent: ISACSON, John, P.; Foley & Lardner, Suite 500,  
3000 K Street, N.W., Washington, DC 20007-5109 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,  
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,  
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,  
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,  
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,  
TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG,  
CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— Without international search report and to be republished  
upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance  
Notes on Codes and Abbreviations" appearing at the beginning  
of each regular issue of the PCT Gazette.

WO 01/12819 A2

(54) Title: NOVEL PROTEIN PHOSPHATASES AND DIAGNOSIS AND TREATMENT OF PHOSPHATASE-RELATED DISORDERS

(57) Abstract: The present invention concerns polypeptides, nucleic acids encoding such polypeptides, cells, tissues and animals containing such nucleic acids, antibodies to the polypeptides, assays utilizing the polypeptides, and methods relating to all of the foregoing. Preferably, the polypeptides of the present invention are phosphatases. Through the use of a "motif extraction" bioinformatics script, additional mammalian members of the phosphatase family are herein presented. These phosphatases include MKP-like proteins, a CDC14-like protein, a PTEN-like protein, and myotubularin (MTM)-like proteins. Classification of proteins as new members of established families has proven highly accurate not only in predicting motifs present in the remaining non-catalytic portion of each protein, but also in their regulation, substrates, and signaling pathways.

**BEST AVAILABLE COPY**

5

10

**NOVEL PROTEIN PHOSPHATASES AND DIAGNOSIS AND TREATMENT OF  
PHOSPHATASE-RELATED DISORDERS**

This application claims priority to U.S. application  
15 serial no. 60/149,005, filed August 13, 1999, the entire  
contents of which, including the figures, are hereby  
incorporated by reference.

**Field of the Invention**

20 The present invention relates to polypeptides. In  
particular, the invention concerns phosphatase polypeptides,  
nucleotide sequences encoding the polypeptides, various  
products and assay methods that can be used for identifying  
compounds useful for the diagnosis and treatment of various  
25 phosphatase-related diseases and conditions, for example  
cell proliferative disorders.

**Background of the Invention**

The following description is provided to aid in  
30 understanding the invention but is not admitted to be prior  
art to the invention.

Cellular signal transduction is a fundamental mechanism whereby external stimuli that regulate diverse cellular processes are relayed to the interior of cells. One of the key biochemical mechanisms of signal transduction involves the reversible phosphorylation of proteins by protein kinases, which enables regulation of the activity of mature proteins by altering their structure and function. The best characterized eukaryotic protein kinases phosphorylate proteins on the alcohol moiety of serine, threonine and tyrosine residues. These kinases largely fall into two groups, those specific for phosphorylating serines and threonines, and those specific for phosphorylating tyrosines. The phosphorylation state of a given substrate also is regulated by the protein phosphatases, a class of proteins responsible for removal of the phosphate group added to a given substrate by a protein kinase. The protein phosphatases can also be classified as being specific for either serine/threonine or tyrosine. Protein phosphatases thus are a large family of enzymes that catalyze the dephosphorylation of proteins modified by phosphorylation of the hydroxyl-containing amino acids serine, threonine or tyrosine. Some members of this family are able to dephosphorylate only tyrosine (the protein tyrosine phosphatases), whereas others are able to dephosphorylate tyrosine as well as serine and threonine (dual-specificity phosphatases). These proteins share a 250-300 amino acid domain that comprises the common catalytic core structure. Related phosphatases are clustered into distinct subfamilies of tyrosine phosphatases, dual-specificity phosphatases, and myotubularin-like phosphatases (Fauman EB, et al., Trends

Biochem Sci. 1996 Nov;21(11):413-7; Martell KJ, et al., Mol Cells. 1998 Feb 28;8(1):2-11).

Through the use of a "motif extraction" bioinformatics script, we have identified additional mammalian members of the phosphatase family. We present here the partial or complete sequence of 20 new phosphatases, their classification, predicted or deduced protein structure, and a strategy for elucidating their biologic and therapeutic relevance. These inventive proteins include 15 MKP-like proteins, two CDC14-like proteins, and two myotubularin (MTM)-like proteins. A PTEN-like protein also is described. Classification of novel proteins as new members of established families has proven highly accurate not only in predicting motifs present in the remaining non-catalytic portion of each protein, but also in their regulation, substrates, and signaling pathways.

Phosphatases have been implicated as regulating a variety of cellular responses, including response to growth factors, cytokines and hormones, oxidative-, UV-, or irradiation-related stress pathways, inflammatory signals (i.e. TNF), apoptotic stimuli (i.e. Fas), T and B cell costimulation, the control of cytoskeletal architecture, and cellular transformation (see The Protein Phosphatase Factsbook, Nick Tonks, Shirish Shenolikar, Harry Charbonneau, Academic Pr, 2000)

Phosphatases also possess a variety of non-catalytic domains that are believed to interact with upstream regulators. Examples include proline-rich domains for interaction with SH3-containing proteins, or specific domains for interaction with Rac, Rho, and Rab small G-

proteins. These interactions may provide a mechanism for cross-talk between distinct biochemical pathways in response to external stimuli such as the activation of a variety of cell surface receptors, including tyrosine kinases, cytokine  
5 receptors, TNF receptor, Fas, T cell receptors, CD28, or CD40.

#### Summary of the Invention

The present invention relates to polypeptides, nucleic  
10 acids encoding such polypeptides, vectors, cells, tissues and animals containing such nucleic acids, antibodies to the polypeptides, assays utilizing the polypeptides, and methods relating to all of the foregoing. Preferably, the polypeptides of the present invention are phosphatases.  
15 Through the use of a "motif extraction" bioinformatics script, additional mammalian members of the phosphatase family are herein presented. These phosphatases include MKP-like proteins, CDC14-like proteins, PTEN-like proteins, and myotubularin (MTM)-like proteins. Classification of  
20 proteins as new members of established families has proven highly accurate not only in predicting motifs present in the remaining non-catalytic portion of each protein, but also in their regulation, substrates, and signaling pathways.

An aspect of the invention features isolated, enriched,  
25 or purified nucleic acid molecules encoding polypeptides, preferably phosphatases. In preferred embodiments, the invention includes an isolated, enriched or purified nucleic acid molecule encoding a phosphatase, wherein said nucleic acid molecule comprises a nucleotide sequence that

(a) encodes a polypeptide having the amino acid sequence set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:42, SEQ ID NO:38 or SEQ ID NO:40;

(b) is the complement of the nucleotide sequence of (a);

10 (c) hybridizes under highly stringent conditions to the molecule of (b) and encodes a naturally occurring polypeptide;

(d) encodes a polypeptide having the full length amino acid sequence set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:38, SEQ ID NO:40 or SEQ ID NO:42, except that it lacks at least one or more, but not all, of the contiguous set of numbered amino acid residues as set forth in the respective domain delimitations in any of the Figures;

(e) is the complement of the nucleotide sequence of (d);

25 (f) the amino acid sequence set forth in at least one of the respective sets of numbered amino acid residues set forth in any Figure;

(g) is the complement of the nucleotide sequence of (f);

(h) encodes a polypeptide having the full length amino acid sequence set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:38, SEQ ID NO:40 or SEQ ID NO:42, except that it lacks one or more, but not all, of the domains selected from the group consisting of an N-terminal domain, a phosphatase domain and a C-terminal domain;

10 (i) is the complement of the nucleotide sequence of (h);

(j) has the nucleotide sequence set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:41, SEQ ID NO:37 or SEQ ID NO:39; or

(k) is the complement of the nucleotide sequence set forth in (j). Preferably, the nucleic acid is isolated, purified or enriched from a mammal, most preferably from a human.

According to another aspect of the present invention, there are provided methods of treating diseases or disorders by administering to a patient an agent that modulates activity of a phosphatase having an amino acid according to the the present invention, such as those identified in the attached figures in view of the teachings contained herein. Due to the broad functional implications of various phosphatase families, such treatment may be effectuated to a wide range of diseases, including cancer, pathophysiological

hypoxia, cardiovascular disorders, Papillon-Lefevre syndrome, Cowden disease, ectodermal dysplasia, Moebius syndrome, Bjornstad syndrome, Bannayan Zonana syndrome, schizophrenia and hamartomas. Of particular significance is  
5 treatment to various type of cancers, as exemplified in Example 3. The method of the present invention may be used to treat breast cancer, urogenital cancer, prostate cancer, head and neck cancer, lung cancer, synovial sarcomas, renal cell carcinoma, non-small cell lung cancer, hepatocellular  
10 carcinoma, pancreatic endocrine tumors, stomach cancer, glioblastoma, colorectal cancer, and thyroid cancer.

The relevance of a phosphatase gene to a particular disease condition can be evaluated in order to effect treatment. According to one embodiment of the present  
15 invention, microarray expression analysis is performed to establish expression profiles of various phosphatase genes according to the invention, and thereby identify the ones whose expression correlates with certain diseased conditions.

20 It should be appreciated that many ways of comparison and correlation analysis may be carried out based on expression data generated in the way similar to that described in Example 3, which become apparent to one skilled in the art based on the above discussion and which therefore  
25 fall in the scope of the invention. Inferences derived from those comparison and correlation analysis may similarly be used in substantiating the treatment method according to this invention. One scenario to be noted is when pairs of samples of normal tissues and diseased tissues are used to  
30 make the expression arrays, the data generated will



specifically demonstrate which phosphatase genes are differentially expressed in certain diseased conditions, thereby form targets of the treatment method according to the present invention. That is, modulators or agents that  
5 are capable of regulating their activities, either in vivo or in vitro, may be identified and used in the treatment of the given diseased conditions.

According to the present invention, there also are provided methods for detection of a phosphatase in a sample  
10 as a diagnostic tool for a disease or disorder using nucleotide probes derived from the phosphatase gene sequences disclosed in the present invention, such as those disclosed herein. Due to the broad functional implications of various phosphatase families, such diagnostic measures  
15 may be used for a wide range of diseases, including cancer, pathophysiological hypoxia, cardiovascular disorders, Papillon-Lefevre syndrome, Cowden disease, ectodermal dysplasia, Moebius syndrome, Bjornstad syndrome, Bannayan Zonana syndrome, schizophrenia and hamartomas. Of  
20 particular importance is diagnose of various type of cancers. The diagnostic method of the present invention may be used to test for breast cancer, urogenital cancer, prostate cancer, head and neck cancer, lung cancer, synovial sarcomas, renal cell carcinoma, non-small cell lung cancer,  
25 hepatocellular carcinoma, pancreatic endocrine tumors, stomach cancer, glioblastoma, colorectal cancer, and thyroid cancer.

Similar to the method of treatment discussed above, it is useful to determine the level of relevance of a  
30 phosphatase gene to a particular diseased condition is

determined in order to effect accurate diagnoses. Such determinations can be accomplished by performing microarray expression analysis according to one embodiment of this invention. The phosphatase genes whose expression  
5 correlates with certain diseased conditions may be identified by the procedure described herein.

Many ways of comparison and correlation analysis may be carried out based on expression data generated in the way similar to that described here; they also necessarily fall  
10 in the scope of the present invention. Inferences derived from those comparison and correlation analysis may similarly be used in substantiating the diagnostic method according to this invention. One scenario to be noted is when pairs of samples of normal tissues and diseased tissues are used to  
15 make the expression arrays, the data generated will specifically demonstrate which phosphatase genes are differentially expressed in certain diseased conditions, therefore may serve as diagnostic markers used in the aforementioned diagnostic method.

20 According to the present invention, there also are provided methods for detection of a phosphatase in a sample as a diagnostic tool for a disease or disorder by comparing a nucleic acid target region of the phosphatase genes disclosed in the present invention, such genes encoding the  
25 amino acid sequences listed in Figure 2, with a control region; and then detecting differences in sequence or amount between the target region and control region as an indication of the disease or disorder. This method also may be used for diagnosing a wide range of diseases, including  
30 cancer, pathophysiological hypoxia, cardiovascular

disorders, Papillon-Lefevre syndrome, Cowden disease, ectodermal dysplasia, Moebius syndrome, Bjornstad syndrome, Bannayan Zonana syndrome, schizophrenia and hamartomas. Of particular importance is diagnose of various type of  
5 cancers. As the aforementioned diagnostic method, this particular method may similarly be used to test for breast cancer, urogenital cancer, prostate cancer, head and neck cancer, lung cancer, synovial sarcomas, renal cell carcinoma, non-small cell lung cancer, hepatocellular  
10 carcinoma, pancreatic endocrine tumors, stomach cancer, glioblastoma, colorectal cancer, and thyroid cancer.

A target region can be any particular region of interest in a phosphatase gene, such as an upstream regulatory region. Variations of sequence in an upstream  
15 regulatory region in a family of phosphatase often have functional implications some of which may be significant in bringing about certain diseased conditions. Changes of the amount of a target region, e.g., changes of number of copies of a regulatory region such as a receptor-binding site, in  
20 certain phosphatase genes, may also represent mechanisms of functional differentiation and hence may be connected to certain diseased states. Detection of such differences in sequence and amount of a target region compared to a control region therefore may effectively lead to detection of a  
25 diseased condition.

In one embodiment of the present invention, microarray studies may be used to identify the potential connections between a diseased condition and variations of a target region among a set of phosphatase genes. For example,  
30 nucleic acid probes may be made that correspond to a given

target region and a control region, respectively, of a phosphatase gene of interest. Samples from normal and diseased tissues are used to make microarray as discussed supra and in Example 3. Hybridization of these probes to  
5 the array so made will yield comparative profiles of the region of interest in the normal and diseased condition, and thus may derive a definition of differences of the target region and control region that is characterized of the disease in question. Such definition in turn may serve as  
10 an indication of the diseased condition as used in the second-mentioned diagnostic method according to the present invention. It should be appreciated that many equivalent or similar methods may be used in carrying out the diagnosis according to the invention which would become apparent to  
15 the skilled person in the art based on the example provided here, and therefore, they are covered in the scope of this invention. The invention is further illuminated by the following explanations.

By "isolated" in reference to a nucleic acid is meant,  
20 for example, a polymer of 14, 17, 21, 35, 50, 75, 100 or more nucleotides conjugated to each other, including DNA or RNA that is isolated from a natural source or that is synthesized. The isolated nucleic acid of the present invention is unique in the sense that it is not found in a  
25 pure or separated state in nature. Use of the term "isolated" indicates that a naturally occurring sequence has been removed from its normal cellular (e.g., chromosomal) environment. Thus, the sequence may be in a cell-free solution or placed in a different cellular environment. The  
30 term does not imply that the sequence is the only nucleotide

sequence present, but that it is substantially free (about 90 - 95% pure at least) of non-nucleotide material naturally associated with it and thus is meant to be distinguished from isolated chromosomes.

5 By the use of the term "enriched" in reference to nucleic acid is meant that the specific DNA or RNA sequence constitutes a significantly higher fraction (2 - 5 fold) of the total DNA or RNA present in the cells or solution of interest than in normal or diseased cells or in the cells  
10 from which the sequence was taken. This could be caused by a person or device by preferential reduction in the amount of other DNA or RNA present, or by a preferential increase in the amount of the specific DNA or RNA sequence, or by a combination of the two. However, it should be noted that  
15 "enriched" does not necessarily imply that there are no other DNA or RNA sequences present, just that the relative amount of the sequence of interest has been significantly increased. The term "significant" here is used to indicate that the level of increase is useful to the person making  
20 such an increase, and generally means an increase relative to other nucleic acids of about at least 2 fold, more preferably at least 5 to 10 fold or even more. The term also does not imply that there is no DNA or RNA from other sources. The other source DNA may, for example, comprise  
25 DNA from a yeast or bacterial genome, or a cloning vector such as pUC19. This term distinguishes the sequence from naturally occurring enrichment events, such as viral infection, or tumor type growths, in which the level of one mRNA may be naturally increased relative to other species of  
30 mRNA. That is, the term is meant to cover only those

situations in which a person has intervened to elevate the proportion of the desired nucleic acid.

It is also advantageous for some purposes that a nucleotide sequence be in purified form. The term  
5 "purified" in reference to nucleic acid does not require absolute purity (such as a homogeneous preparation); instead, it represents an indication that the sequence is relatively purer than in the natural environment (compared to the natural level this level should be at least 2-5 fold  
10 greater, e.g., in terms of mg/ml). Individual clones isolated from a cDNA library may be purified to electrophoretic homogeneity. The claimed DNA molecules obtained from these clones can be obtained directly from total DNA or from total RNA. The cDNA clones are not  
15 naturally occurring, but rather are preferably obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The construction of a cDNA library from mRNA involves the creation of a synthetic substance (cDNA) and pure individual cDNA clones can be  
20 isolated from the synthetic library by clonal selection of the cells carrying the cDNA library. Thus, the process which includes the construction of a cDNA library from mRNA and isolation of distinct cDNA clones preferably yields more than approximately 100 fold purification and more preferably  
25 yields an approximately 10<sup>6</sup>-fold purification of the native message. Thus, purification of at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated. The term is also chosen to distinguish clones  
30 already in existence which may encode phosphatases but which

have not been isolated from other clones in a library of clones. Thus, the term covers clones encoding phosphatases which are isolated from other non-phosphatase clones.

Nucleic acids that hybridize to any of the above  
5 sequences or functional derivatives (as defined below) of any of the above are also contemplated as part of the invention. The nucleic acid may be isolated from a natural source by cDNA cloning, enrichment hybridization and/or subtractive hybridization techniques; the natural source may  
10 be mammalian (human) blood, semen, or tissue and the nucleic acid may be synthesized by the triester or other method or by using an automated DNA synthesizer.

The term "hybridize" refers to a method of interacting a nucleic acid sequence with a DNA or RNA molecule in  
15 solution or on a solid support, such as cellulose or nitrocellulose. If a nucleic acid sequence binds to the DNA or RNA molecule with high affinity, it is said to "hybridize" to the DNA or RNA molecule. The strength of the interaction between the probing sequence and its target can  
20 be assessed by varying the stringency of the hybridization conditions. Various low or high stringency hybridization conditions may be used depending upon the specificity and selectivity desired (see for example, Berger et al., Methods in Enzymology, Guide to Molecular Cloning Techniques Volume  
25 152 (1987), the entire content of which is hereby incorporated by reference in its entirety, including any drawings). Stringency is controlled by varying salt or denaturant concentrations. By high stringent hybridization assay conditions is meant hybridization assay conditions at  
30 least as stringent as the following: hybridization in 50%

formamide, 5X SSC, 50 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 6.8, 0.5% SDS, 0.1 mg/mL sonicated salmon sperm DNA, and 5X Denhart solution at 42 °C overnight; washing with 2X SSC, 0.1% SDS at 45 °C; and washing with 0.2X SSC, 0.1% SDS at 45 °C. More highly stringent conditions include 0.1X SSC, 0.05% SDS and 55 °C for the second wash.

One skilled in the art will recognize how such conditions can be altered to vary specificity and selectivity. Under highly stringent hybridization conditions only highly complementary nucleic acid sequences hybridize. Preferably, such conditions prevent hybridization of nucleic acids having one or two mismatches out of 20 contiguous nucleotides.

In yet other preferred embodiments the nucleic acid is an isolated conserved or unique region, for example those useful for the design of hybridization probes to facilitate identification and cloning of additional polypeptides, or for the design of PCR probes to facilitate cloning of additional polypeptides.

By "conserved nucleic acid regions", it is meant regions present on two or more nucleic acids encoding a polypeptide, preferably a phosphatase polypeptide, to which a particular nucleic acid sequence can hybridize under lower stringency conditions. Examples of lower stringency conditions suitable for screening for nucleic acids encoding phosphatase polypeptides are provided in Abe, et al. J. Biol. Chem. 19:13361 (1992); and Berger et al., above (hereby incorporated by reference herein in its entirety, including any drawings). Preferably, conserved regions differ by no more than 5 out of 20 contiguous nucleotides.



By "unique nucleic acid region" it is meant a sequence present in a full length nucleic acid coding for a phosphatase polypeptide that is not present in a sequence coding for any other known naturally occurring polypeptide.

5 Such regions preferably comprise 14, 17, 21, 35, 50, 75, 100 or more contiguous nucleotides present in the full length nucleic acid encoding a phosphatase polypeptide. In particular, a unique nucleic acid region is preferably of human origin. A unique nucleic acid region may be

10 identified by aligning the full length sequence of interest with a previously known sequence. The two sequences will each have a number of contiguous nucleotides that are identical to one another. A unique nucleic acid region will contain these contiguous nucleotides and one or more

15 additional contiguous nucleotides from the full length sequence of interest.

The invention also features a nucleic acid probe for the detection of a nucleic acid encoding a phosphatase polypeptide in a sample. The nucleic acid probe contains

20 nucleic acid that will hybridize specifically to a sequence of, for example, at least 14, 17, 21, 35, 50, 75, 100 or more contiguous nucleotides set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID

25 NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:41, SEQ ID NO:37 or SEQ ID NO:39 or a complement or a functional derivative thereof. The probe is preferably at least 14, 17, 21, 35, 50, 75, 100 or more bases in length

and selected to hybridize specifically to a unique region of a phosphatase encoding nucleic acid.

In preferred embodiments the nucleic acid probe hybridizes to a nucleic acid encoding a polypeptide having the amino acid sequence set forth in at least one of the respective sets of numbered amino acid residues set forth in any Figure. Various low or high stringency hybridization conditions may be used depending upon the specificity and selectivity desired, as recited above. Under highly stringent hybridization conditions only highly complementary nucleic acid sequences hybridize. Preferably, such conditions prevent hybridization of nucleic acids having 1 or 2 mismatches out of 20 contiguous nucleotides.

Methods for using the probes include detecting the presence or amount of phosphatase RNA in a sample by contacting the sample with a nucleic acid probe under conditions such that hybridization occurs and detecting the presence or amount of the probe bound to phosphatase RNA. The nucleic acid duplex formed between the probe and a nucleic acid sequence coding for a phosphatase polypeptide may be used in the identification of the sequence of the nucleic acid detected (for example see, Nelson et al., in Nonisotopic DNA Probe Techniques, p. 275 Academic Press, San Diego (Kricka, ed., 1992) hereby incorporated by reference herein in its entirety, including any drawings). Kits for performing such methods may be constructed to include a container means having disposed therein a nucleic acid probe.

The invention also features recombinant nucleic acid, preferably in a cell or an organism. The recombinant

nucleic acid may contain a sequence encoding any of the phosphatases set forth, or functional derivatives thereof, and a vector or a promoter effective to initiate transcription in a host cell. The recombinant nucleic acid  
5 can alternatively contain a transcriptional initiation region functional in a cell, a sequence complimentary to an RNA sequence encoding a phosphatase polypeptide and a transcriptional termination region functional in a cell.

Another aspect of the invention features an isolated,  
10 enriched or purified polypeptide. Preferably, the isolated, enriched or purified polypeptide is a phosphatase polypeptide. The polypeptide of the present invention comprises an amino acid sequence having

(a) the amino acid sequence set forth in SEQ ID NO:2,  
15 SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:42, SEQ ID NO:38 or SEQ ID NO:40;

20 (b) the amino acid sequence set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:38, SEQ ID NO:40 or SEQ ID  
25 NO:42, except that it lacks one or more, but not all, of the respective domain delimitations set forth in any of the Figures;

(c) the amino acid sequence set forth in at least one of the respective sets of numbered amino acid residues set  
30 forth in any Figure;

(d) the amino acid sequence set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32 or SEQ ID NO:34, SEQ ID NO:38, SEQ ID NO:40 or SEQ ID NO:42, except that it lacks at least one, but not all, of the following domains: an N-terminal domain, a C-terminal domain or a phosphatase domain. Preferably, the phosphatase is isolated, purified or enriched from a mammal, most preferably from a human.

By "phosphatase polypeptide" it is meant an amino acid sequence substantially similar to the sequence shown in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:42, SEQ ID NO:38 or SEQ ID NO:40, or fragments thereof. A sequence that is substantially similar will preferably have at least 90% identity (more preferably at least 95% and most preferably 99-100%) to the sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:42, SEQ ID NO:38 or SEQ ID NO:40.

By "identity" is meant a property of sequences that measures their similarity or relationship. Identity is measured by dividing the number of identical residues in the two sequences by the total number of residues and multiplying the product by 100. Thus, two copies of exactly

the same sequence have 100% identity, but sequences that are less highly conserved and have deletions, additions, or replacements have a lower degree of identity. Those skilled in the art will recognize that several computer programs are available for determining sequence identity. Using standard parameters, for example Gapped BLAST or PSI-BLAST (Altschul, et al. (1997) Nucleic Acids Res. 25:3389-3402), BLAST (Altschul, et al. (1990) J. Mol. Biol. 215:403-410), and Smith-Waterman (Smith, et al. (1981) J. Mol. Biol. 147:195-197).

By "isolated" in reference to a polypeptide is meant, for example, a polymer of 6, 12, 18, 24, 30, 36, 50, 75, 100 or more amino acids conjugated to each other, including polypeptides that are isolated from a natural source or that are synthesized. The isolated polypeptides of the present invention are unique in the sense that they are not found in a pure or separated state in nature. Use of the term "isolated" indicates that a naturally occurring sequence has been removed from its normal cellular environment. Thus, the sequence may be in a cell-free solution or placed in a different cellular environment. The term does not imply that the sequence is the only amino acid chain present, but that it is essentially free (about 90 - 95% pure at least) of material naturally associated with it.

By the use of the term "enriched" in reference to a polypeptide it is meant that the specific amino acid sequence constitutes a significantly higher fraction (2 - 5 fold) of the total of amino acids present in the cells or solution of interest than in normal or diseased cells or in the cells from which the sequence was taken. This could be

caused by a person by preferential reduction in the amount of other amino acids present, or by a preferential increase in the amount of the specific amino acid sequence of interest, or by a combination of the two. However, it should be noted that "enriched" does not imply that there are no other amino acid sequences present, just that the relative amount of the sequence of interest has been significantly increased. The term significant here is used to indicate that the level of increase is useful to the person making such an increase, and generally means an increase relative to other amino acids of about at least 2 fold, more preferably at least 5 to 10 fold or even more. The term also does not imply that there is no amino acid from other sources. The other amino acid may, for example, comprise amino acid encoded by a yeast or bacterial genome, or a cloning vector such as pUC19. The term is meant to cover only those situations in which a person has intervened to elevate the proportion of the desired nucleic acid.

It is also advantageous for some purposes that an amino acid sequence be in purified form. The term "purified" in reference to a polypeptide does not require absolute purity (such as a homogeneous preparation); instead, it represents an indication that the sequence is relatively purer than in the natural environment (compared to the natural level this level should be at least 2-5 fold greater, e.g., in terms of mg/ml). Purification of at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated. The substance is preferably free of contamination at a

functionally significant level, for example at least 90%, 95%, or 99% pure.

In another aspect the invention features an isolated, enriched, or purified polypeptide fragment, preferably a phosphatase polypeptide fragment. By "a phosphatase polypeptide fragment" it is meant an amino acid sequence that is less than the full-length phosphatase amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:42, SEQ ID NO:38 or SEQ ID NO:40. Examples of fragments include phosphatase domains, phosphatase mutants and phosphatase-specific epitopes or recombinant phosphatase polypeptide.

By a "domain" it is meant a portion of the polypeptide having homology to one or more known proteins wherein the sequence predicts some common function, interaction or activity. Polypeptide domains of the present invention include C-terminal domains, N-terminal domains and phosphatase domains (i.e., the catalytic domain as provided, alternatively, throughout the disclosure).

The term "phosphatase domain" refers to the region of the protein phosphatase that is responsible for excising a phosphate from a phosphorylated protein.

The term "N-terminal domain" refers to the extracatalytic region located between the initiator methionine, or first amino acid if the N-terminal domain is partial, and the catalytic domain of the protein phosphatase. The N-terminal domain can be identified

following a Smith-Waterman alignment of the protein sequence against the non-redundant protein database to define the N-terminal boundary of the catalytic domain. Depending on its length, the N-terminal domain may or may not play a  
5 regulatory role in phosphatase function.

The C-terminal domain refers to the region located between the catalytic domain and the carboxy-terminal amino acid residue of the phosphatase. The C-terminal domain can be identified following a Smith-Waterman alignment of the  
10 protein sequence against the non-redundant protein database to define the C-terminal boundary of the catalytic domain. The C-terminal domain may or may not play a regulatory role in phosphatase function. In the present invention, either the N-terminal or C-terminal domain may encompass  
15 unidentified domains responsible for additional protein function.

By a "phosphatase mutant" it is meant a phosphatase polypeptide which differs from the native sequence in that one or more amino acids have been changed, added and/or  
20 deleted. Changes in amino acids may be conservative or non-conservative. By "conservative" it is meant the substitution of an amino acid for one with similar properties such as charge, hydrophobicity, structure, etc. Examples of polypeptides encompassed by this term include,  
25 but are not limited to, (1) chimeric proteins which comprise a portion of a phosphatase polypeptide sequence fused to a non-phosphatase polypeptide sequence, for example a polypeptide sequence of hemagglutinin (HA), (2) phosphatase proteins lacking a specific domain, for example the  
30 catalytic domain, and (3) phosphatase proteins having a



point mutation. A phosphatase mutant will retain some useful function such as, for example, binding to a natural binding partner, catalytic activity, or the ability to bind to a phosphatase-specific antibody (as defined below).

5 By "phosphatase-specific epitope" it is meant a sequence of amino acids that is both antigenic and unique to a phosphatase. Phosphatase-specific epitopes can be used to produce phosphatase-specific antibodies, as more fully described below.

10 By "recombinant phosphatase polypeptide" it is meant to include a polypeptide produced by recombinant DNA techniques such that it is distinct from a naturally occurring polypeptide either in its location (e.g., present in a different cell or tissue than found in nature), purity or  
15 structure. Generally, such a recombinant polypeptide will be present in a cell in an amount different from that normally observed in nature.

Yet another aspect of the invention features an antibody (e.g., a monoclonal or polyclonal antibody) having  
20 specific binding affinity to a polypeptide or polypeptide fragment, the polypeptide preferably being a phosphatase. By "specific binding affinity" is meant that the antibody binds to phosphatase polypeptides with greater affinity than it binds to other polypeptides under specified conditions.

25 The antibody of the present invention has specific binding affinity to a phosphatase or a fragment thereof, wherein said phosphatase or fragment thereof has the amino acid sequence set forth in at least one of the respective sets of numbered amino acid residues set forth in any Figure.

Antibodies having specific binding affinity to a phosphatase polypeptide may be used in methods for detecting the presence and/or amount of a phosphatase polypeptide in a sample by contacting the sample with the antibody under  
5 conditions such that an immunocomplex forms and detecting the presence and/or amount of the antibody conjugated to the phosphatase polypeptide. Diagnostic kits for performing such methods may be constructed to include a first container containing the antibody and a second container having a  
10 conjugate of a binding partner of the antibody and a label, such as, for example, a radioisotope. The diagnostic kit may also include notification of an FDA approved use and instructions therefor.

In another aspect the invention features a hybridoma  
15 which produces an antibody having specific binding affinity to a polypeptide of the present invention. By "hybridoma" is meant an immortalized cell line which is capable of secreting an antibody, for example a phosphatase antibody. In preferred embodiments the phosphatase antibody comprises  
20 a sequence of amino acids that is able to specifically bind the phosphatase molecules of the present invention.

In another embodiment, the invention encompasses a recombinant cell or tissue containing a purified nucleic acid coding for a polypeptide, preferably a phosphatase  
25 polypeptide. The recombinant cell of the present invention comprises a nucleic acid molecule, wherein said nucleic acid molecule encodes a phosphatase having the the amino acid sequence set forth in at least one of the respective sets of numbered amino acid residues set forth in any Figure or a  
30 functional equivalent thereof. In such cells, the nucleic

acid may be under the control of its genomic regulatory elements, or may be under the control of exogenous regulatory elements including an exogenous promoter. By "exogenous" it is meant a promoter that is not normally coupled transcriptionally to the coding sequence for the phosphatase polypeptide in its native state.

The invention features a method for identifying human cells containing a polypeptide or a related sequence. The method involves identifying the polypeptide in human cells using techniques that are routine and standard in the art, such as those described herein for identifying phosphatase (e.g., cloning, Southern or Northern blot analysis, in situ hybridization, PCR amplification, etc.).

The invention also features methods of screening cells for natural binding partners of polypeptides. By "natural binding partner" it is meant a protein that interacts with a polypeptide, preferably a phosphatase. Binding partners include ligands, agonists, antagonists and downstream signaling molecules such as adaptor proteins and may be identified by techniques well known in the art such as co-immunoprecipitation or by using, for example, a two-hybrid screen. (Fields and Song, U.S. Patent No. 5,283,173, issued February 1, 1994 and, incorporated by reference herein.) The present invention also features the purified, isolated or enriched versions of the polypeptides identified by the methods described above.

In another aspect, the invention provides an assay to identify substances that modulate the activity of a polypeptide, preferably a phosphatase, comprising the steps of

(a) contacting at least one phosphatase having the amino acid sequence set forth in at least one of the respective numbered amino acid residues as set forth in any Figure;

- 5       (b) measuring an activity of the phosphatase; and  
      (c) determining whether the test substance modulates the activity of the phosphatase.

Such assays may be performed in vitro or in vivo and can be obtained by modifying existing assays, such as the  
10 assays described in WO 96/40276, published December 19, 1996 and WO 96/14433, published May 17, 1996 (both incorporated herein by reference including any drawings). Other possibilities include testing for phosphatase activity on standard substrates. The substances so identified may be  
15 enhancers or inhibitors of phosphatase activity and can be peptides, natural products (such as those isolated from fungal strains, for example) or small molecular weight chemical compounds. A preferred substance will be a compound with a molecular weight of less than 5,000, more  
20 preferably less than 1,000, most preferably less than 500. The assay and substances contemplated by the invention are discussed in more detail below.

Another aspect of the invention is a method for identifying a substance that modulates a phosphatase  
25 activity in a cell comprising the steps of

- (a) expressing at least one phosphatase having the amino acid sequence set forth in at least one of the respective numbered amino acid residues as set forth in any Figure;  
30       (b) adding a test substance to the cell; and

(c) monitoring

(i) a change in cell phenotype or

(ii) the interaction between the phosphatase and natural binding partner.

- 5 For example, inhibitors of phosphatase activity can be tested as treatments for cell proliferative disorders such as leukemia or lymphoma using subcutaneous xenograph models in mice.

10 In another aspect of the invention, a method for treating a disease or disorder is provided comprising the step of administering to a patient in need of such a treatment a substance that modulates an activity of a polypeptide, preferably a phosphatase, having the amino acid sequence set forth in at least one of the respective  
15 numbered amino acid residues as set forth in any Figure. The disease or disorder may be cancer, pathophysiological hypoxia such as seen in cardiac dysfunction and vascular disorders including atherosclerosis, stenosis and stroke, myopathies, congenital muscle disorders, Papillon-Lefevre  
20 syndrome, Cowden disease, ectodermal dysplasia, Moebius syndrome, Bjornstad syndrome, Bannayan-Zonana syndrome, glioblastoma, schizophrenia and hamartomas. The cancer may be breast cancer, glioblastoma, urogenital cancer, prostate cancer, head and neck cancer, lung cancer, synovial  
25 sarcomas, renal cell carcinoma, non-small cell lung cancer, hepatocellular carcinoma, pancreatic endocrine tumors, stomach cancer, colorectal cancer and thyroid cancer. Phosphatase activity may be stimulated and the method may modulate activity in vitro.

In yet another aspect of the invention, a method for detection of a polypeptide, preferably a phosphatase, in a sample as a diagnostic tool for a disease or disorder is provided, comprising the steps of

5 (a) contacting said sample with a nucleic acid probe which hybridizes under hybridization assay conditions to a nucleic acid which encodes a polypeptide having the amino acid sequence set forth in at least one of the respective sets of numbered amino acid residues set forth in any

10 Figure; and

(b) detecting the presence or amount of a probe:target region as an indication of the disease. The disease or disorder may be cancer, pathophysiological hypoxia such as seen in cardiac disfunction and vascular disorders including  
15 atherosclerosis, stenosis and stroke, myopathies, congenital muscle disorders, Papillon-Lefevre syndrome, Cowden disease, ectodermal dysplasia, Moebius syndrome, Bjornstad syndrome, Bannayan-Zonana syndrome, glioblastoma, schizophrenia and hamartomas. The cancer may be breast cancer, glioblastoma,  
20 urogenital cancer, prostate cancer, head and neck cancer, lung cancer, synovial sarcomas, renal cell carcinoma, non-small cell lung cancer, hepatocellular carcinoma, pancreatic endocrine tumors, stomach cancer, colorectal cancer and thyroid cancer.

25 In yet another aspect of the invention, a method for detection of a polypeptide, preferably a phosphatase, in a sample as a diagnostic tool for a disease or disorder, wherein said method comprises the steps of

(a) comparing a nucleic acid target region, said  
30 nucleic acid encoding said polypeptide, in a sample to a

control region, wherein said polypeptide has the amino acid sequence set forth in at least one of the respective sets of numbered amino acid residues set forth in any Figure; and

(b) detecting differences in sequence or sequence amount between said target region and said control region as an indication of the disease or disorder. The disease or disorder may be cancer, pathophysiological hypoxia such as seen in cardiac disfunction and vascular disorders including atherosclerosis, stenosis and stroke, myopathies, congenital muscle disorders, Papillon-Lefevre syndrome, Cowden disease, ectodermal dysplasia, Moebius syndrome, Bjornstad syndrome, Bannayan-Zonana syndrome, glioblastoma, schizophrenia and hamartomas. The cancer may be breast cancer, glioblastoma, urogenital cancer, prostate cancer, head and neck cancer, lung cancer, synovial sarcomas, renal cell carcinoma, non-small cell lung cancer, hepatocellular carcinoma, pancreatic endocrine tumors, stomach cancer, colorectal cancer and thyroid cancer.

The summary of the invention described above is non-limiting and other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

#### **Brief Description of the Figures**

Figure 1 shows a comparison of phosphatase polynucleotides of the invention. From left to right, each of the columns stands, respectively, for: a & b) the designation given each sequence by the inventors, c) the species of the isolated sequence, d & e) the SEQ ID NO: of the nucleotide and amino acid sequence, respectively, f)

whether the ORF is full-length (FL) or encodes a full-length catalytic domain (CAT), g), h) and i) the Super-Family, Group and Family of the phosphatase, as defined by the text of the specification, j) and k) sequence length in

- 5 nucleotides and amino acids, respectively, l-n) ORF start and end and ORF length, respectively, o-r) DNA repeats, SNP position, chromosomal localization and whether "Expression" patterns are set forth in the specification;

Figure 2 shows a comparison of phosphatases of the invention. From left to right, columns a) through e) are identical to those in Figure 1, above. Columns f) and g) recite the Group and Family, respectively, of the genes. Columns h) and i) identify the start and end of catalytic domains. Columns j) and k) recite the start and end of rhodanase domains. Columns l) through q) recite nraa  
15 Pscore, the length, in contiguous amino acids, over which the match is determined, the ID match, the identity percentage, the similarity percentage and the nraa match ACC#, respectively, of the nucleic acids of the invention. Finally, the last column gives a brief description of the  
20 gene and/or amino acid of the invention;

Figure 3 shows results from expression profiles. From left to right, columns a) and b) indicate the sample and source. Column c) identifies the tag, and d) the cell type. Column e) provides illustrative comments. Columns f) through  
25 k) respectively identify the tumor-sym, normal-sym, tumor-lo, tumor cells, normal and p53. Column l) provides activity values. Columns m) through s) respectively set forth data obtain from using the following sequences:

30 SEQ\_ID\_11\_AA374753, SEQ\_ID\_21\_AA915932, SEQ\_ID\_27\_AI031656,



SEQ\_ID\_31\_NP\_060746 (G77-8-14), SEQ\_ID\_33\_NP\_060232  
(AA232384), SEQ\_ID\_37\_MTMR7 (AA663875), and  
SEQ\_ID\_39\_AA493915. See Figure 4.

Figure 4 shows nucleotide sequences according to the  
5 invention; and

Figure 5 shows amino acid sequences according to the  
invention.

#### Detailed Description of the Invention

10 The present invention relates to the isolation and  
characterization of new polypeptides, nucleotide sequences  
encoding these polypeptides, various products and assay  
methods that can be used to identify compounds useful for  
the diagnosis and treatment of various polypeptide-related  
15 diseases and conditions, for example cancer. Polypeptides,  
preferably phosphatases, and nucleic acids encoding such  
polypeptides may be produced using well-known and standard  
synthesis techniques when given the sequences presented  
herein.

20 The polypeptides described in the present invention  
belong to the dual-specificity group of protein  
phosphatases. This classification employs on the conserved  
core amino acid sequence motifs that make up the catalytic  
domain of this class of phosphatases. The unique signature  
25 motifs of the catalytic domain of the dual-specificity class  
of phosphatases is responsible for the ability of these  
enzymes to dephosphorylate phosphoserine/phosphothreonine as  
well phosphotyrosine residues.

The dual-specificity group of protein phosphatases is  
30 divided into family members that include the Cdc14

phosphatases, MAP kinase phosphatases (MKP), myotubularins (MTM), a subclass of the MTM represented by Sbf1 that act as anti-phosphatases, low molecular weight (LMW) Cdc25-like phosphatases and PTEN, the lipid phosphatases. On the basis  
5 of sequence homology, the phosphatases featured in the present invention belong to the following distinct families of dual-specificity phosphatases: Cdc14, MKP4, MTM, MTM-like SBF1 class of antiphosphatases and PTEN. A description  
10 of the structural and functional characteristics for the known family prototypes, together with a summary of the closest homologs of each phosphatase taken from data presented in Figure 3 is presented below.

#### Cdc14 family

15 The Cdc14 family of dual-specific phosphatases is named after its founding member, the Cdc14 phosphatase from *Saccharomyces cerevisiae*, an enzyme that plays a key role in the regulation of the mitotic exit pathway of the cell cycle. Two mammalian Cdc14 phosphatases are known, Cdc14A1  
20 (AF064102) and Cdc14B1 (AF064105). The catalytic domains of these phosphatases (138 and 135 amino acids in length, respectively) exhibit 72% sequence identity. Flanking the catalytic region are relatively long N-terminal (195 and 230 amino acids in Cdc14A1 and Cdc14B1, respectively) and C-  
25 terminal domains (291 and 132 amino acids in Cdc14A1 and Cdc14B1, respectively). The N-terminal domain of Cdc14 appears to be highly conserved between human as well as distant evolutionary homologs, showing 65%, 41% and 35% sequence identity over 178, 193 and 193 amino acids to human  
30 Cdc14B1, *C. elegans* (U28739) and yeast (Q00684) Cdc14,

respectively. The C-terminal domain of Cdc14 exhibits a lower level of sequence conservation than the N-terminal domain, i.e. 35% identity over 151 amino acids between human Cdc14A1 and Cdc14B1 and no detectable homology to the equivalent regions from yeast and *C. elegans* Cdc14. The functional significance of the highly conserved N-terminal domain of Cdc14 is unknown but it is possible that this domain participates in protein-protein interactions that regulate the activity as well as subcellular distribution of Cdc14 as described next.

Studies in yeast have implicated Cdc14 in the regulation of the mitotic exit pathway through inactivation of cyclin-dependent kinases (CDKs) (Visintin R, et al. Mol Cell. (1998) 2:709-718). Cdc14 is part of a nucleolar protein complex called RENT (regulator of nucleolar silencing and telophase) that contains the proteins Net1 (also called Cfi) and Sir2. From G1 through anaphase Cdc14 is found sequestered in the nucleolus where its enzymatic activity is inhibited by Net1. In late anaphase Cdc14 dissociates from the RENT complex in a step dependent on the GTPase Tem1 (Shou W. et al. (1999) Cell 16: 233-44). The released Cdc14 relocates to the nucleus where it inactivates the mitotic CDKs.

Cdc14-mediated CDK inactivation is believed to occur via two mechanisms. First, the ubiquitin-dependent proteolytic system APC (anaphase-promoting complex) degrades the mitotic Clb cyclin normally required for CDK activity. Cdc14 stimulates Clb breakdown by dephosphorylating the APC-specific factor Cdh1.2 (also called Hct1). Second, the kinase inhibitor Sic1 binds to CDKs inhibiting their

activity. Cdc14 activates Sic1 transcription by dephosphorylating the Sic1 transcription factor Swi5. In addition, Cdc14 enhances the stability of the Sic1 protein by dephosphorylating it (Visintin, R. et al. (1999) Nature 398:818-23).

The 154 amino acid partial murine AA023073 (SEQ ID. NO:2) is thought to be closest to Z83760, a putative open reading frame (ORF) from the ascidian organism *ciona intestinalis*, with 68% identity over 59 amino acids corresponding to the catalytic domain. The next closest homolog to AA023073 is believed to be human Cdc14B2 (AF064105) with 40% identity over 65 amino acids. AA023073 has the conserved features of a dual-specificity phosphatase including a catalytic cysteine.

A human orthologue of the murine AA023073 exists in the form of the EST AF086553 with 94% amino acid sequence identity to AA023073 over 115 amino acids. The putative phosphatase encoded by AA023073 is predicted to be a catalytically active phosphatase. Based on its homology to Cdc14, AA023073 may function in mitotic regulation.

#### MKP family

The MAP kinase phosphatases (MKP) family of dual-specificity phosphatases define an important class of enzymes that play a pivotal role in negative feedback regulation of the MAP kinase pathway. This family of phosphatases has 11 family members and we describe herein additional homologs. Included within the known MKPs are DUS1 (also known as MPK-1, CL100, PTPN-10, erp, VH1 or 3CH134), DUS3 (also known as VHR), DUS4 (also known as HVH2, TYP1, MKP2 or VH2), DUS5

(also known as HVH3, B23, VH3). DUS6 (also known as PYST1, MKP3, rVH6), DUS7 (also known as PYST2). CDKN3 (also known as CDKN3, KAP, CIP2 or CDI1), VH5 and STYX.

Structurally MKPs consist of two domains, an N- and a  
5 C-terminal catalytic domain. The N-terminal domain ranges in  
size from about 147 to about 206 amino acids and exhibits  
limited homology (28-47%) among the various family members.  
The N-terminal region of MKP's features two pockets of  
homology termed CH2 domains as well as a 126 amino acid  
10 rhodanase-like motif; these features are conserved with the  
Cdc25 phosphatase and serve an unknown function. The  
catalytic domain of the MKP family members varies in size  
from about 147 to about 206 amino acids and displays 40-74%  
amino acid sequence identity.

15 Most MKP phosphatases are capable of inactivating  
through a dephosphorylation reaction kinases that  
participate in the MAPK pathways. The ERK (extracellular  
signal-regulated kinase), JNK/SAPK (c-Jun N-terminal  
kinase/stress-activated protein kinase) and p38 MAP kinase  
20 pathways mediate the signal transduction events that are  
responsible for cell division, differentiation or apoptosis  
in response to extracellular ligands (Cobb M.H., Prog  
Biophys. Mol. Biol. (1999) 71:479-500). Full MAP kinase  
enzymatic activation requires the concomitant  
25 phosphorylation by selective upstream dual-specificity  
kinases of threonine and tyrosine residues residing in the  
activation loop of the MAP kinases. MKP family dual-  
specificity phosphatases mediate MAP kinase inactivation by  
dephosphorylating these threonine and tyrosine residues.

This mechanism provides negative feedback regulation to the MAP kinase pathways.

MKPs may play a significant role in human cancer by attenuating MAP kinase cascades involved in cellular transformation.

Given the large number of MAP kinases as well as MKPs, a central question is whether there is selectivity in kinase substrate recognition by MKPs. Evidence that such specificity exists is provided by DUS-6 (MKP-3) and VH5 which have been shown to be highly selective phosphatases towards the ERK or JNK/SAPK and p38 MAP kinases, respectively (Muda M, et al., J. Biol. Chem. (1996) 271:27205-8). Another level of substrate specificity comes from subcellular compartmentalization as shown by DUS-6 (MKP-3) which is found exclusively in the cytosol rather than in the nucleus (Groom, L.A. et al. (1996) EMBO J. 15: 3621-3632). Further selectivity can arise at the level of the tissue specificity of expression (Muda, M. et al. (1997) J. Biol. Chem. 272:5141-5151).

MKPs appear to be as ubiquitous in their phylogenetic distribution as their MAP kinase counterparts with multiple members present in yeast ( e.g. YVH1), C. elegans ( e.g. Y042), Drosophila, ( e.g. puckered ), plants ( e.g. DsPTP1) and mammals. The primary mode of action of MKPs isolated from different species appears to be MAPK dephosphorylation thereby providing negative feedback to the MAPK signal transduction pathways.

MKPs may play an important role during pathophysiological hypoxia as suggested by the induction of MKP-1 gene expression under low oxygen conditions

(Laderoute, K. R. (1999) J. Biol. Chem. 274:12890-12897). Tumor hypoxia is directly linked to the onset of angiogenesis during malignant progression (Hanahan, D. et al. (1996) Cell 86:353-364 and Mazure, N.M. et al. (1996) 5 Cancer Res. 56:3436-3440). A number of genes have been found to be induced during hypoxic conditions such as the heat shock transcription factor-1 (HSF-1) (Benjamin, I.J. et al. (1990) Proc. Natl. Acad. Sci. 87:6263-6267), c-fos and c-jun (Ausserer, W.A. et al. (1994) Mol. Cell. Biol. 14:5032-5042, 10 and Muller, J.M. (1997) J. Biol. Chem 272:23435-23439) and the hypoxia-inducible factor-1 (HIF-1) (Wenger, R.H. et al. (1997) J. Biol. Chem. 378:609-616). MKP-1 transcripts and protein have been shown to be upregulated in early-stage carcinomas well as in multiple stages of breast and prostate 15 carcinomas ( e.g. Leav, I. et al. (1996) Lab. Invest. 75: 361-370). The role of enhanced MKP-1 expression in cancer has not been elucidated. Since hypoxic conditions are known to trigger apoptosis via the activation of the JNK pathway (reviewed in Ip, Y.T. et al. (1998) Curr. Opin. Cell Biol. 20 10:205-219) and MAPK phosphatases provide negative feedback to this pathway, it is conceivable that MKP-1 supports tumor growth by blocking apoptosis. The dephosphorylation and subsequent inactivation of ERK-1 and ERK-2 by MAPK phosphatases may also be responsible for suppressing 25 angiogenic vascular endothelial cell proliferation by angiostatin ( Redlitz, A. et al. (1999) J. Vasc. Res 36:28-34).

The MKP phosphatases of the present invention may have as their primary function negative feedback regulation of 30 MAPK signal transduction. Since there is precedence for

selectivity in the mechanism of action at the level of substrate recognition, subcellular localization and tissue distribution among the known MKPs, the MKPs described may display similar selectivity. The MKPs may also play a role  
5 in suppressing apoptosis by blocking the JNK/SAPK pathway during pathological hypoxia such as that occurring in angiogenic tumors. The development of specific phosphatase inhibitors that target the anti-apoptotic MKPs may prove valuable as an approach to cancer therapy.

10 The 176 amino acid human protein coded for by SGP033 (AA Seq ID#26) is 50% identical to the known human dual specificity phosphatase MKP1-like (NP\_008957). It has two regions of repeat DNA (323-341, 541-559, Figure 1). SGP033 (AA Seq ID#26) has been mapped to human chromosomal region  
15 2q33-q37.2.

The 163 amino acid murine protein AA030322 (AA Seq ID#04) is 31% identical to the human MKP1-like phosphatase NP\_008957. It is the murine orthologue to human SGP033 (AA Seq ID#26), with 80% identity in 158 aa overlap. AA030322  
20 (NA Seq ID#03) contains a repeat region at nucleotide position 95-114.

The 184 amino acid full length human gene AA374753 (AA SEQ ID#12) shares 56% amino acid identity to the dual specificity phosphatase CG10089, a gene product from  
25 *Drosophila melanogaster*. This gene is expressed in fetal brain, testis and thymus. The 184 amino acid gene AA103595 (AA SEQ ID #6) is the murine orthologue of human AA374753 (AA SEQ ID #12) with 94% identity over 184 amino acids.

The 198 amino acid full length human LOC51207 (AA SEQ ID#18) is closely related to the public sequence DUS13  
30



protein phosphatase NP\_057448 [Homo sapiens], with 99% identity over 198 amino acids. LOC51207 maps to human chromosome 10q21.3. The 198 amino acid full length murine AA144705 (AA SEQ ID#08) is the murine orthologue to the 198 amino acid full-length human LOC51207 (AA SEQ ID#18) with 88% identity over 198 amino acids.

The 217 amino acid full length human AI031656 (AA SEQ ID#28) is closest to the predicted ORF AAF67187, MAP kinase phosphatase-1 [Drosophila melanogaster], with 41% identity over 78 amino acids. The expression is higher in tumor samples than in normal samples. The 220 amino acid murine AA274457 (AA SEQ ID#10) is closely related to the 217 amino acid human AI031656 (AA SEQ ID# 28) with 84% identity over 212 amino acids.

NA SEQ ID#29 maps to chromosome 1q32.1 and has a repeat region between nucleotides 102-152. The 218 amino acid murine AA396428 (AA SEQ ID#14) is closest to the 482 amino acid human MKP5 (AA SEQ ID#30) with 95% identity over 154 amino acids. It is the murine MKP5.

The 340 amino acid human YVH1 (AA923158, AA SEQ ID#24) is identical to the 340 amino acid public sequence NP\_009171. This gene maps to chromosomal position 1q21-q22. The 339 amino acid murine AA422661 (AA SEQ ID#16) is closest to the 340 amino acid human YVH1 (AA SEQ ID# 24) with 84% identity over 339 amino acids.

The 190 amino acid human gene AA813123 (AA SeqID#20) is 95% identical over 190 amino acids to the public sequence AAD33910, MKP-like protein phosphatase [Homo sapiens]. AA813123 (NA SeqID#19) has two repeat regions (11-28; 187-

204), and the gene maps to human chromosomal position Xp11.4-q12.

The 188 amino acid human gene AA915932 (AA Seq ID#22) is 50% identical to the public sequence NP\_008957, MKP-1 like [Homo sapiens]. The DNA sequence (NA Seq ID#21) has a repeat region at position 410-427, and the gene maps to human chromosome 22q12.1-qter.

NP\_060746 has a repeat region at sequence 915-938; it maps to human chromosome 11q12-q13.2 and is expressed at a high level in fetal brain and testis.

#### MTM family

MTMs have been shown to be capable of dephosphorylating phosphoserine and phosphotyrosine residues (Laporte, J. et al. (1998) Human Molecular Genetics, 7:1703-1712). Structurally MTMs consist of a central 200 amino acid catalytic region flanked by N-terminal and C-terminal domains that range in size between 250-400 and 50-300, respectively, among the mammalian, yeast and elegans MTMs. Sbf1 contains a similar domain structure as MTM except that its 200 amino acid central MTM catalytic-like region is flanked by a much longer N-terminal domain (1160 amino acids) and by a similar size C-terminal domain (335 amino acids). Among MTM family members, including Sbf1, the N- and C-terminal domains conserve two and one, pockets of homology, respectively. The functional role of the conserved regions is unknown. The lack of a predicted transmembrane domain in any of the MTMs as well as in Sbf1 suggests that these proteins localize to and function within the

intracellular environment. The tissue distribution of all the known human MTMs is ubiquitous except for MTMR7 appears to be confined to brain (Laporte, J. et al. (1998) Hum. Mol. Genetics 7:1703-1712).

5 In contrast, an important subclass of the MTM family of dual-specificity phosphatases represented by Sbf1 is enzymatically inactive and may function biologically as an anti-phosphatase, an activity which may be responsible for its oncogenic potential (Cui, X. et al. (1998) Nature  
10 Genetics 18:331-337). Sbf1 lacks the conserved HCSDGW signature motif required for catalysis having instead the sequence GLEDGW. In addition, Sbf1 contains a helix-turn-helix region located between 68-92 residues C-terminal to the GLEDGW motif. This motif defines the SID (SET protein-  
15 interaction domain) domain which mediates the interaction between Sbf1 and SET-binding factors such as the proto-oncogene Hrx, the mammalian homologue of drosophila trithorax (Trx). SET (Suvar3-9, Enhancer-of zeste, Tri-  
20 trithorax)-binding factors such as human Hrx and drosophila Trx and enhancer of zeste are proteins that participate in gene regulation (Cui, X. et al. (1998) Nature Genetics 18:331-337). The mechanism of anti-phosphatase action by Sbf1 may involve direct competition with MTMs for substrates or, alternatively, this protein may function as an adaptor  
25 molecule with affinity for phosphorylated proteins.

The classical prototype of the MTM family of dual-specificity phosphatases is MTM1 (myotubularin). Mutations in the MTM1 gene (Xq27-q28) are responsible for X-linked myotubular myopathy (XLMTM) (OMIM 310400,  
30 <http://www.ncbi.nlm.nih.gov/Omim/searchomim.html>), a severe

congenital muscle disorder characterized by hypotonia and respiratory insufficiency that results in high neonatal mortality. MTM1 is conserved from yeast [i.e. scMTMH in *S. cerevisiae* (Z49610)] to mammals with 8 MTM family members identified in humans (Laporte J, et al. Hum Mol Genet. 1998 Oct;7(11):1703-12). The pathological consequences of MTM mutations may not be limited to MTM1 since other human MTM genes are located in chromosomal loci associated with a wide range of conditions. For example, human MTMR2 (11q22) maps within the locus for Papillon-Lefevre syndrome (PLS) (OMIM 245000), a syndrome associated with premature periodontal destruction of the teeth; human MTMR6 (13q12) is a candidate for ectodermal dysplasia (OMIM 129500) and Moebius syndrome (OMIM 157900). In addition, studies with murine syntenic counterparts of various human MTM genes reveal additional potential disease association for this class of phosphatases. Human MTMR3 corresponds to the mouse mutants belted and dilution-peru, human MTMR5 to gray tremor (gt) and human MTMR7 to disorganization (ds) and wobbler-lethal (wl) (Laporte, J. et al (1998) Human Mol. Genetics 7: 1703-1712).

There is growing evidence that mutations in MTM1 genes are associated with disease. The pathological consequences of MTM mutations may not be limited to MTM1 since other human MTM genes are located in chromosomal loci associated with a wide range of conditions. For example, human MTMR2 (11q22) maps within the locus for Papillon-Lefevre syndrome (PLS) (OMIM 245000), a syndrome associated with premature periodontal destruction of the teeth; human MTMR6 (13q12) is a candidate for ectodermal dysplasia (OMIM 129500) and

Moebius syndrome (OMIM 157900). In addition, studies with murine syntenic counterparts of various human MTM genes reveal additional potential disease associations for this class of phosphatases. Human MTMR3 corresponds to the mouse mutants belted (bt) and dilution-peru (dp), human MTMR5 to gray tremor (gt) and human MTMR7 to disorganization (ds) and wobbler-lethal (wl) (Laporte, J. et al. (1998) Human Mol. Genetics 7: 1703-1712).

Two of the MTM-like genes described of the present invention, human AA232238 and AA251929, belong to the Sbf1 MTM-like family of anti-phosphatases. The potential ORFs encoded by these genes lack the canonical HCS motif required for catalytic activity in dual-specificity phosphatases, yet they display homology to MTM's as summarized below. In addition, AA232384 and AA251929 may possess a SID domain at an equivalent position as found in Sbf1 that may participate in binding to SET domain proteins. The third MTM-like gene, human AA663875, lacks a catalytic domain but bears strong homology to the N-terminal domain of the MTM-like protein (CAB38778.1) that contains the catalytic HCS signature motif. Hence, AA663875 is predicted to encode an enzymatically active MTM-like phosphatase.

The 400 amino acid full-length human AA23238 (SEQ ID NO: 34) is thought to be closest to MTM6 (AF072928) and MTM1 (AF002223) with 43 and 42% sequence identity over 114 and 126 amino acids, respectively. The 52 amino acid partial human AA251929 (SEQ ID NO: 36) is believed to be closest to human MTM3 (U58034) with 49% identity over 41 amino acids.

The 138 amino acid partial human AA663875 (SEQ ID NO: 38) is thought to be closest to an MTM-like ORF predicted

from human chromosome X at q11.2-12 (CAB38778.1) with 48% identity over 89 amino acids.

The MTM phosphatase and MTM-like antiphosphatases described in the present invention are likely to play  
5 central roles in the regulation of signalling pathways involved in cell differentiation, cell division and apoptosis.

#### PTEN family

10 The tumor suppressor PTEN (also known as MMAC1 or TEP1) is the prototypical member of the PTEN family of a new and unusual class of enzymes that act as lipid phosphatases. PTEN was first discovered as a tumor suppressor gene isolated from human glioblastomas (Li, J. et al. (1997)  
15 Science 275:1943-1947) and named after recognizing its homology to tensin and auxillin. The PTEN gene (10q23) is mutated in patients with Cowden disease (CD) (OMIM 158350) and Bannayan-Zonana syndrome (153480), conditions characterized by multiple hamartomas. CD patients are at  
20 increased risk of developing malignancies of multiple tissue origins including breast, urogenital, digestive and thyroid (e.g. Nelen, M.R. (1999) Europ. J. Hum Genet. 7:267-73).

PTEN expression inhibits the growth and tumorigenicity of human glioblastoma cells (Li, D.M. et al. (1998) Proc.  
25 Nat. Acad. Sci. 95: 15406-15411). The growth suppression activity of PTEN is mediated by its ability to block cell cycle progression in the G1 phase. PTEN modulates G1 cell cycle progression through negative regulation of the PI3-kinase (PI3K)/Akt pathway by dephosphorylating the  
30 phospholipid phosphatidylinositol (3,4,5)-triphosphate

(PIP3), the activator of AKT. Down-modulation of the AKT pathway leads to increased levels of the universal cyclin-dependent kinase (CDK) inhibitor p27 (KIP1) and, consequently, inhibition of CDK activity and G1 arrest (Li, D.M. et al. (1998) Proc. Nat. Acad. Sci. 95: 15406-15411).

The ability of PTEN to associate with and dephosphorylate the focal adhesion kinase (FAK), both in vivo and in vitro, suggests that PTEN may also down-modulate FAK-induced signalling events related to cell spreading and motility (Tamura, M. et al. (1998) Science 280: 1614-1617). The PTEN/FAK interaction has been recently linked to suppression of the phospholipid-activated PI3K/Akt cell survival pathway (Tamura, M. et al. (1999) 274:20693-20703). These findings strongly suggest that PTEN may play an important role in the processes of cell invasion and metastasis in cancer.

PTEN is phylogenetically conserved having close homologs in yeast (YNL128W), C. elegans (daf-18, CAA10315) (Mihaylova V. T. et al. (1999) Proc Natl Acad Sci. 96:7427-32) and mammals. In C. elegans, DAF-18 acts as a negative regulator of the DAF-2 and AGE-1 (PI3K/Akt) signaling pathway, consistent with the notion that DAF-18 acts a phosphatidylinositol 3,4,5-triphosphate phosphatase in vivo.

Two human PTEN genes are known, PTEN1 (U92436) and PTEN2 (AF01083). These encode proteins that are 403 amino acids long and 98% identical over their entire length and consist of a central 156 amino acid catalytic domain flanked by 23 and 224 amino acid N- and C-terminal domains, respectively. The C-terminus of PTEN has a PDZ domain that may be involved in important protein-protein interactions at

membrane or cytoskeletal interfaces. The exact function of the extended C-terminal domain is presently unknown although it is conceivable that this domain plays an important role in localizing PTEN to the proper membrane/cytoskeletal environment where phosphoinositide synthesis occurs. At this site the balance between PI3 kinase/PTEN activities is likely to determine the effective concentration of diffusable second messengers such as PIP3. Since PIP3 is a potent activator not only of AKT, but of other important signalling proteins such as Vav, PKC, PLC and Btk as well (reviewed in: Maehama, T. and Dixon, J.E. (1999) Trends Cell Biol. 9:125-128), the activity of PTEN as a phosphoinositide phosphatase is likely to play a pivotal role in the attenuation of diverse signalling downstream pathways.

The 357 amino acid human AA493915 (AA SEQ ID#40) is 65% identical over 357 amino acids to human TPTE, or "transmembrane phosphatase with tensin homology", NP\_037447. The novel PTEN-like phosphatase represented by human AA493915 (AA SEQ ID#40) may have, like PTEN, tumor suppressor function. AA493915 (AA SEQ ID#40) may prove valuable in signalling events that mediate cell growth and apoptosis as well as in the design of novel therapeutic agents to treat cancer. Expression of AA493915 in many tissues including cerebellum, pituitary, prostate, fetal brain and fetal lung, as shown in Figure 3.

The polypeptide and nucleotide sequences of the invention can be used, therefore, to identify modulators of cell growth and survival which are useful in developing therapeutics for various cell proliferative disorders and conditions, and in particular cancers related to



inappropriate phosphatase activity. Assays to identify compounds that act intracellularly to enhance or inhibit phosphatase activity can be developed by creating genetically engineered cell lines that express phosphatase  
5 nucleotide sequences, as is more fully discussed below.

### **I. Nucleic Acids Encoding Polypeptides.**

An aspect of the invention features nucleic acid sequences encoding a polypeptide. Included within the scope  
10 of this invention are the functional equivalents of the herein-described isolated nucleic acid molecules. Functional equivalents or derivatives can be obtained in several ways. The degeneracy of the genetic code permits substitution of certain codons by other codons which specify  
15 the same amino acid and hence would give rise to the same protein. The nucleic acid sequence can vary substantially since, with the exception of methionine and tryptophan, the known amino acids can be coded for by more than one codon. Thus, portions or all of the polypeptide genes could be  
20 synthesized to give a nucleic acid sequence significantly different from that shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID  
25 NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:41, SEQ ID NO:37 or SEQ ID NO:39. The encoded amino acid sequence thereof would, however, be preserved.

In addition, the nucleic acid sequence may comprise a nucleotide sequence which results from the addition,  
30 deletion or substitution of at least one nucleotide to the

5'-end and/or the 3'-end of the nucleic acid formula shown in any of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:41, SEQ ID NO:37 or SEQ ID NO:39 or a derivative thereof. Any nucleotide or polynucleotide may be used in this regard, provided that its addition, deletion or substitution does not alter the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:42, SEQ ID NO:38 or SEQ ID NO:40 which is encoded by the nucleotide sequence. For example, the present invention is intended to include any nucleic acid sequence resulting from the addition of ATG as an initiation codon at the 5'-end of the nucleic acid sequence or its functional derivative, or from the addition of TTA, TAG or TGA as a termination codon at the 3'-end of the inventive nucleotide sequence or its derivative. Moreover, the nucleic acid molecule of the present invention may, as necessary, have restriction endonuclease recognition sites added to its 5'-end and/or 3'-end.

Such functional alterations of a given nucleic acid sequence afford an opportunity to promote secretion and/or processing of heterologous proteins encoded by foreign nucleic acid sequences fused thereto. All variations of the nucleotide sequence of the phosphatase genes and fragments

thereof permitted by the genetic code are, therefore, included in this invention.

Further, it is possible to delete codons or to substitute one or more codons with codons other than degenerate codons to produce a structurally modified polypeptide which has substantially the same utility or activity of the polypeptide produced by the unmodified nucleic acid molecule. As recognized in the art, the two polypeptides are functionally equivalent, as are the two nucleic acid molecules which give rise to their production, even though the differences between the nucleic acid molecules are not related to degeneracy of the genetic code.

Functional equivalents or derivatives of polypeptides can also be obtained using nucleic acid molecules encoding one or more functional domains of the polypeptide. For example, the catalytic domain of phosphatases function as an enzymatic remover of phosphate molecules bound onto tyrosine and/or serine/threonine amino acids and a nucleic acid sequence encoding the catalytic domain alone or linked to other heterologous nucleic acid sequences can be considered a functional derivative of phosphatases.

## **II. A Nucleic Acid Probe for the Detection of Polypeptides.**

A nucleic acid probe of the present invention may be used to probe an appropriate chromosomal or cDNA library by art recognized hybridization methods to obtain another nucleic acid molecule of the present invention. A chromosomal DNA or cDNA library may be prepared from appropriate cells according to recognized methods in the art (e.g. "Molecular Cloning: A Laboratory Manual", second

edition, edited by Sambrook, Fritsch, & Maniatis, Cold Spring Harbor Laboratory, 1989).

In the alternative, chemical synthesis may be carried out in order to obtain nucleic acid probes having nucleotide sequences which correspond to N-terminal and C-terminal portions of the amino acid sequence of the polypeptide of interest. Thus, the synthesized nucleic acid probes may be used as primers in a polymerase chain reaction (PCR) carried out in accordance with recognized PCR techniques, essentially according to PCR Protocols, "A Guide to Methods and Applications", edited by Michael et al., Academic Press, 1990, utilizing the appropriate chromosomal or cDNA library to obtain the fragment of the present invention.

One skilled in the art can readily design such probes based on the sequence disclosed herein using methods of computer alignment and sequence analysis known in the art (e.g.. "Molecular Cloning: A Laboratory Manual", second edition, edited by Sambrook, Fritsch, & Maniatis, Cold Spring Harbor Laboratory, 1989). The hybridization probes of the present invention can be labeled by standard labeling techniques such as with a radiolabel, enzyme label, fluorescent label, biotin-avidin label, chemiluminescence, and the like. After hybridization, the probes may be visualized using known methods.

The nucleic acid probes of the present invention include RNA as well as DNA probes and nucleic acids modified in the sugar phosphate or even the base portion as long as the probe still retains the ability to specifically hybridize under conditions as disclosed herein. Such probes are generated using techniques known in the art. The

nucleic acid probe may be immobilized on a solid support. Examples of such solid supports include, but are not limited to, plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, acrylic resins, such as polyacrylamide and latex beads, and nitrocellulose. Techniques for coupling nucleic acid probes to such solid supports are well known in the art.

The test samples suitable for nucleic acid probing methods of the present invention include, for example, cells or nucleic acid extracts of cells, or biological fluids. The sample used in the above-described methods will vary based on the assay format, the detection method and the nature of the tissues, cells or extracts to be assayed. Methods for preparing nucleic acid extracts of cells are well known in the art and can be readily adapted in order to obtain a sample which is compatible with the method utilized.

### **III. A Probe Based Method And Kit For Detecting Polypeptides.**

One method of detecting the presence of polypeptides in a sample comprises (a) contacting the sample with the above-described nucleic acid probe, under conditions such that hybridization occurs, and (b) detecting the presence of the probe bound to the nucleic acid molecule. One skilled in the art would select the nucleic acid probe according to techniques known in the art as described above. Samples to be tested include but should not be limited to RNA samples of human tissue. In preferred embodiments, high stringency hybridization conditions are used.

A kit for detecting the presence of a polypeptide in a sample comprises at least one container having disposed therein the above-described nucleic acid probe. The kit may further comprise other containers comprising one or more of the following: wash reagents and reagents capable of detecting the presence of bound nucleic acid probe. Examples of detection reagents include, but are not limited, to radiolabelled probes, enzymatically labeled probes (e.g. horseradish peroxidase, alkaline phosphatase), and affinity labeled probes (e.g. biotin, avidin, or streptavidin).

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allow the efficient transfer of reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the probe or primers used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, and the like), and containers which contain the reagents used to detect the hybridized probe, bound antibody, amplified product, or the like. One skilled in the art will readily recognize that the nucleic acid probes described in the present invention can readily be incorporated into one of the established kit formats which are well known in the art with or without a

set of instructions concerning the use of such reagents in an assay.

#### IV. DNA Constructs Comprising a Nucleic Acid

##### 5           Molecule and Cells Containing These Constructs.

The present invention also relates to a recombinant DNA molecule comprising, 5' to 3', a promoter effective to initiate transcription in a host cell and the above-described nucleic acid molecules. In addition, the present  
10 invention relates to a recombinant DNA molecule comprising a vector and a nucleic acid molecule described herein. The present invention also relates to a nucleic acid molecule comprising a transcriptional region functional in a cell, a sequence complimentary to an RNA sequence encoding an amino  
15 acid sequence corresponding to a polypeptide or functional derivative, and a transcriptional termination region functional in said cell. The above-described molecules may be isolated and/or purified DNA molecules.

The present invention also relates to a cell or  
20 organism that contains a nucleic acid molecule as described herein and thereby is capable of expressing a peptide. The polypeptide may be purified from cells which have been altered to express the polypeptide. A cell is said to be "altered to express a desired polypeptide" when the cell,  
25 through genetic manipulation, is made to produce a protein which it normally does not produce or which the cell normally produces at lower levels. One skilled in the art can readily adapt procedures for introducing and expressing either genomic, cDNA, or synthetic sequences into either  
30 eukaryotic or prokaryotic cells.

A nucleic acid molecule, such as DNA, is said to be "capable of expressing" a polypeptide if it contains nucleotide sequences which contain transcriptional and translational regulatory information and such sequences are "operably linked" to nucleotide sequences which encode the polypeptide. An operable linkage is a linkage in which the regulatory DNA sequences and the DNA sequence sought to be expressed are connected in such a way as to permit gene sequence expression. The precise nature of the regulatory regions needed for gene sequence expression may vary from organism to organism, but will in general include a promoter region which, in prokaryotes, contains both the promoter (which directs the initiation of RNA transcription) as well as the DNA sequences which, when transcribed into RNA, will signal synthesis initiation. Such regions will normally include those 5'-non-coding sequences involved with initiation of transcription and translation, such as the TATA box, capping sequence, CAAT sequence, and the like.

If desired, the non-coding region 3' to the sequence encoding a polypeptide gene may be obtained by the above-described cloning methods. This region may be retained for its transcriptional termination regulatory sequences, such as termination and polyadenylation. Thus, by retaining the 3'-region naturally contiguous to the DNA sequence encoding a polypeptide gene, the transcriptional termination signals may be provided. Where the transcriptional termination signals are not satisfactorily functional in the expression host cell, then a 3' region functional in the host cell may be substituted.



Two DNA sequences (such as a promoter region sequence and a phosphatase sequence) are said to be operably linked if the nature of the linkage between the two DNA sequences does not (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the promoter region sequence to direct the transcription of a phosphatase gene sequence, or (3) interfere with the ability of the phosphatase gene sequence to be transcribed by the promoter region sequence. Thus, a promoter region would be operably linked to a DNA sequence if the promoter were capable of effecting transcription of that DNA sequence. Thus, to express a phosphatase gene, transcriptional and translational signals recognized by an appropriate host are necessary.

The present invention encompasses the expression of a gene (or a functional derivative thereof) in either prokaryotic or eukaryotic cells. Prokaryotic hosts are, generally, very efficient and convenient for the production of recombinant proteins and are, therefore, one type of preferred expression system for a gene. Prokaryotes most frequently are represented by various strains of *E. coli*. However, other microbial strains may also be used, including other bacterial strains.

In prokaryotic systems, plasmid vectors that contain replication sites and control sequences derived from a species compatible with the host may be used. Examples of suitable plasmid vectors may include pBR322, pUC118, pUC119 and the like; suitable phage or bacteriophage vectors may include  $\lambda$ gt10,  $\lambda$ gt11 and the like; and suitable virus vectors may include pMAM-neo, pKRC and the like.

Preferably, the selected vector of the present invention has the capacity to replicate in the selected host cell.

Recognized prokaryotic hosts include bacteria such as *E. coli* and those from genera such as *Bacillus*,  
5 *Streptomyces*, *Pseudomonas*, *Salmonella*, *Serratia*, and the like. However, under such conditions, the polypeptide will not be glycosylated. The prokaryotic host must be compatible with the replicon and control sequences in the expression plasmid.

10 To express a polypeptide (or a functional derivative thereof) in a prokaryotic cell, it is necessary to operably link a polypeptide sequence to a functional prokaryotic promoter. Such promoters may be either constitutive or, more preferably, regulatable (e.g., inducible or  
15 derepressible). Examples of constitutive promoters include the *int* promoter of bacteriophage  $\lambda$ , the *bla* promoter of the  $\beta$ -lactamase gene sequence of pBR322, and the CAT promoter of the chloramphenicol acetyl transferase gene sequence of pPR325, and the like. Examples of inducible prokaryotic  
20 promoters include the major right and left promoters of bacteriophage  $\lambda$  (PL and PR), the *trp*, *recA*, *lacZ*, *lacI*, and *gal* promoters of *E. coli*, the  $\alpha$ -amylase (Ulmanen et al., J. Bacteriol. 162:176-182, 1985) and the  $\zeta$ -28-specific promoters of *B. subtilis* (Gilman et al., Gene sequence  
25 32:11-20(1984)), the promoters of the bacteriophages of *Bacillus* (Gryczan, In: The Molecular Biology of the Bacilli, Academic Press, Inc., NY (1982)), and *Streptomyces* promoters (Ward et al., Mol. Gen. Genet. 203:468-478, 1986). Prokaryotic promoters are reviewed by Glick (J. Ind.  
30 Microbiot. 1:277-282, 1987); Cenatiempo (Biochimie 68:505-

516, 1986); and Gottesman (Ann. Rev. Genet. 18:415-442, 1984).

Proper expression in a prokaryotic cell also requires the presence of a ribosome-binding site upstream of the gene sequence-encoding sequence. Such ribosome binding sites are disclosed, for example, by Gold et al. (Ann. Rev. Microbiol. 35:365-404, 1981). The selection of control sequences, expression vectors, transformation methods, and the like, are dependent on the type of host cell used to express the gene.

As used herein, "cell", "cell line", and "cell culture" may be used interchangeably and all such designations include progeny. Thus, the words "transformants" or "transformed cells" include the primary subject cell and cultures derived therefrom, without regard to the number of transfers. It is also understood that all progeny may not be precisely identical in DNA content, due to deliberate or inadvertent mutations. However, as defined, mutant progeny have the same functionality as that of the originally transformed cell.

Host cells which may be used in the expression systems of the present invention are not strictly limited, provided that they are suitable for use in the expression of the peptide of interest. Suitable hosts may often include eukaryotic cells. Preferred eukaryotic hosts include, for example, yeast, fungi, insect cells, mammalian cells either in vivo, or in tissue culture. Mammalian cells which may be useful as hosts include HeLa cells, cells of fibroblast origin such as VERO, 3T3 or CHO-K1, or cells of lymphoid origin (such as 32D cells) and their derivatives. Preferred

mammalian host cells include SP2/0 and J558L, as well as neuroblastoma cell lines such as IMR 332 and PC12 which may provide better capacities for correct post-translational processing.

5        In addition, plant cells are also available as hosts, and control sequences compatible with plant cells are available, such as the cauliflower mosaic virus 35S and 19S, and nopaline synthase promoter and polyadenylation signal sequences. Another preferred host is an insect cell, for  
10        example the *Drosophila* larvae. Using insect cells as hosts, the *Drosophila* alcohol dehydrogenase promoter can be used. Rubin, Science 240:1453-1459, 1988). Alternatively, baculovirus vectors can be engineered to express large amounts of a phosphatase in insects cells (Jasny, Science  
15        238:1653, 1987); Miller et al., In: Genetic Engineering (1986), Setlow, J.K., et al., eds., Plenum, Vol. 8, pp. 277-297).

Any of a series of yeast gene sequence expression systems can be utilized which incorporate promoter and  
20        termination elements from the actively expressed gene sequences coding for glycolytic enzymes which are produced in large quantities when yeast are grown in mediums rich in glucose. Known glycolytic gene sequences can also provide very efficient transcriptional control signals. Yeast  
25        provides substantial advantages in that it can also carry out post-translational peptide modifications. A number of recombinant DNA strategies exist which utilize strong promoter sequences and high copy number of plasmids which can be utilized for production of the desired proteins in  
30        yeast. Yeast recognizes leader sequences on cloned

mammalian gene sequence products and secretes peptides bearing leader sequences ( e.g., pre-peptides). For a mammalian host, several possible vector systems are available for the expression of a phosphatase.

5       A particularly preferred yeast expression system is that utilizing *Schizosaccharomyces pombe*. This system is useful for studying the activity of members of the Src family (Superti-Furga, et al. EMBO J. 12:2625, 1993) and other NR-TKs.

10       A wide variety of transcriptional and translational regulatory sequences may be employed, depending upon the nature of the host. The transcriptional and translational regulatory signals may be derived from viral sources, such as adenovirus, bovine papilloma virus, cytomegalovirus,  
15       simian virus, or the like, where the regulatory signals are associated with a particular gene sequence which has a high level of expression. Alternatively, promoters from mammalian expression products, such as actin, collagen, myosin, and the like, may be employed. Transcriptional  
20       initiation regulatory signals may be selected which allow for repression or activation, so that expression of the gene sequences can be modulated. Of interest are regulatory signals which are temperature-sensitive so that by varying the temperature, expression can be repressed or initiated,  
25       or are subject to chemical (such as metabolite) regulation.

      Expression of polypeptide in eukaryotic hosts requires the use of eukaryotic regulatory regions. Such regions will, in general, include a promoter region sufficient to direct the initiation of RNA synthesis. Preferred  
30       eukaryotic promoters include, for example, the promoter of

the mouse metallothionein I gene sequence (Hamer et al., J. Mol. Appl. Gen. 1:273-288, 1982); the TK promoter of Herpes virus (McKnight, Cell 31:355-365, 1982); the SV40 early promoter (Benoist et al., Nature (London) 290:304-310, 1981); the yeast gal4 gene sequence promoter (Johnston et al., Proc. Natl. Acad. Sci. (USA) 79:6971-6975, 1982); Silver et al., Proc. Natl. Acad. Sci. (USA) 81:5951-5955, 1984).

Translation of eukaryotic mRNA is initiated at the codon which encodes the first methionine. For this reason, it is preferable to ensure that the linkage between a eukaryotic promoter and a DNA sequence which encodes a polypeptide (or a functional derivative thereof) does not contain any intervening codons which are capable of encoding a methionine (e.g., AUG). The presence of such codons results either in a formation of a fusion protein (if the AUG codon is in the same reading frame as a coding sequence) or a frame-shift mutation (if the AUG codon is not in the same reading frame as a coding sequence).

A polypeptide nucleic acid molecule and an operably linked promoter may be introduced into a recipient prokaryotic or eukaryotic cell either as a nonreplicating DNA (or RNA) molecule, which may either be a linear molecule or, more preferably, a closed covalent circular molecule (a plasmid). Since such molecules are incapable of autonomous replication, the expression of the gene may occur through the transient expression of the introduced sequence. Alternatively, permanent or stable expression may occur through the integration of the introduced DNA sequence into the host chromosome.

A vector may be employed which is capable of integrating the desired gene sequences into the host cell chromosome. Cells which have stably integrated the introduced DNA into their chromosomes can be selected by  
5 also introducing one or more markers which allow for selection of host cells which contain the expression vector. The marker may provide for prototrophy to an auxotrophic host, biocide resistance, e.g., antibiotics, or heavy metals, such as copper, or the like. The selectable marker  
10 gene sequence can either be directly linked to the DNA gene sequences to be expressed, or introduced into the same cell by co-transfection. Additional elements may also be needed for optimal synthesis of single chain binding protein mRNA. These elements may include splice signals, as well as  
15 transcription promoters, enhancers, and termination signals. cDNA expression vectors incorporating such elements include those described by Okayama, Mol. Cell. Bio. 3:280, 1983.

The introduced nucleic acid molecule can be incorporated into a plasmid or viral vector capable of  
20 autonomous replication in the recipient host. Any of a wide variety of vectors may be employed for this purpose. Factors of importance in selecting a particular plasmid or viral vector include: the ease with which recipient cells that contain the vector may be recognized and selected from  
25 those recipient cells which do not contain the vector; the number of copies of the vector which are desired in a particular host; and whether it is desirable to be able to "shuttle" the vector between host cells of different species.

Preferred prokaryotic vectors include plasmids such as those capable of replication in *E. coli* (such as, for example, pBR322, ColE1, pSC101, pACYC 184,  $\pi$ VX. Such plasmids are, for example, disclosed by Sambrook (cf.

5 "Molecular Cloning: A Laboratory Manual", second edition, edited by Sambrook, Fritsch, & Maniatis, Cold Spring Harbor Laboratory, (1989)). *Bacillus* plasmids include pC194, pC221, pT127, and the like. Such plasmids are disclosed by Gryczan (In: *The Molecular Biology of the Bacilli*, Academic  
10 Press, NY (1982), pp. 307-329). Suitable *Streptomyces* plasmids include p1J101 (Kendall et al., *J. Bacteriol.* 169:4177-4183, 1987), and *streptomyces* bacteriophages such as  $\phi$ C31 (Chater et al., In: *Sixth International Symposium on Actinomycetales Biology*, Akademiai Kiado, Budapest, Hungary  
15 (1986), pp. 45-54). *Pseudomonas* plasmids are reviewed by John et al. (*Rev. Infect. Dis.* 8:693-704, 1986), and Izaki (*Jpn. J. Bacteriol.* 33:729-742, 1978).

Preferred eukaryotic plasmids include, for example, BPV, vaccinia, SV40, 2-micron circle, and the like, or their  
20 derivatives. Such plasmids are well known in the art (Botstein et al., *Miami Wntr. Symp.* 19:265-274, 1982); Broach, In: *"The Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance"*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, p. 445-470  
25 (1981); Broach, *Cell* 28:203-204, 1982); Bollon et al., *J. Clin. Hematol. Oncol.* 10:39-48, 1980); Maniatis, In: *Cell Biology: A Comprehensive Treatise*, Vol. 3, Gene Sequence Expression, Academic Press, NY, pp. 563-608 (1980).

Once the vector or nucleic acid molecule containing the  
30 construct(s) has been prepared for expression, the DNA



construct(s) may be introduced into an appropriate host cell by any of a variety of suitable means, e.g., transformation, transfection, conjugation, protoplast fusion, electroporation, particle gun technology, calcium phosphate-precipitation, direct microinjection, and the like. After the introduction of the vector, recipient cells are grown in a selective medium, which selects for the growth of vector-containing cells. Expression of the cloned gene molecule(s) results in the production of polypeptides or fragments or functional derivatives thereof. This can take place in the transformed cells as such, or following the induction of these cells to differentiate (for example, by administration of bromodeoxyuracil to neuroblastoma cells or the like). A variety of incubation conditions can be used to form the peptide of the present invention. The most preferred conditions are those which mimic physiological conditions.

#### V. Polypeptides

Also a feature of the invention are polypeptides, preferably phosphatases. A variety of methodologies known in the art can be utilized to obtain the polypeptides of the present invention. They may be purified from tissues or cells which naturally produce them. Alternatively, the above-described isolated nucleic acid sequences can be used to express a protein recombinantly.

Any eukaryotic organism can be used as a source for the polypeptide of the invention, as long as the source organism naturally contains such a polypeptide. As used herein, "source organism" refers to the original organism from which the amino acid sequence is derived, regardless of the

organism the protein is expressed in and ultimately isolated from.

One skilled in the art can readily follow known methods for isolating proteins in order to obtain the peptide free of natural contaminants. These include, but are not limited to: size-exclusion chromatography, HPLC, ion-exchange chromatography, and immuno-affinity chromatography.

A phosphatase protein, like all proteins, is comprised of distinct functional units or domains. In eukaryotes, proteins sorted through the so-called vesicular pathway (bulk flow) usually have a signal sequence (also called a leader peptide) in the N- terminus, which is cleaved off after the translocation through the ER (endoplasmic reticulum) membrane. Some N-terminal signal sequences are not cleaved off, remaining as transmembrane segments, but it does not mean these proteins are retained in the ER; they can be further sorted and included in vesicles. Non-receptor proteins generally function to transmit signals within the cell, either by providing sites for protein:protein interactions or by having some catalytic activity (contained within a catalytic domain), often both. Methods of predicting the existence of these various domains are well known in the art. Protein:protein interaction domains can be identified by comparison to other proteins. The SH2 domain, for example is a protein domain of about 100 amino acids first identified as a conserved sequence region between the proteins Src and Fps (Sadowski, et al., Mol. Cell. Bio. 6:4396, 1986). Similar sequences were later found in many other intracellular signal-transducing proteins. SH2 domains function as regulatory modules of

intracellular signaling cascades by interacting with high affinity to phosphotyrosine-containing proteins in a sequence specific and strictly phosphorylation-dependent manner (Mayer and Baltimore, Trends Cell. Biol. 3:8, 1993).

- 5 Kinase or phosphatase catalytic domains can be identified by comparison to other known catalytic domains with kinase or phosphatase activity. See, for example Hanks and Hunter, FASEB J. 9:576-595, 1995.

- 10 Phosphatase domains have a variety of uses. An example of such a use is to make a polypeptide consisting of the phosphatase catalytic domain and a heterologous protein such as glutathione S-transferase (GST). Such a polypeptide can be used in a biochemical assay for phosphatase catalytic activity useful for studying phosphatase substrate
- 15 specificity or for identifying substances that can modulate phosphatase catalytic activity. Alternatively, one skilled in the art could create a polypeptide lacking at least one of three major domains, a extracellular domain, transmembrane domain or intracellular domain. Such a
- 20 polypeptide, when expressed in a cell, is able to form complexes with the natural binding partner(s) of phosphatases but unable to transmit any signal further downstream into the cell, e.g., it would be signaling incompetent and thus would be useful for studying the
- 25 biological relevance of phosphatase activity. (See, for example, Gishizky, et al, PNAS :10889, 1995).

**VI. An Antibody Having Binding Affinity To A Polypeptide  
And A Hybridoma Containing the Antibody.**

The present invention also relates to an antibody having specific binding affinity to a polypeptide. The polypeptide may have the amino acid sequence set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:42, SEQ ID NO:38 or SEQ ID NO:40, or a be fragment thereof, or at least 6, 12, 18, 24, 30, 36, 50, 75, 100 contiguous amino acids thereof. Such an antibody may be identified by comparing its binding affinity to a first polypeptide with its binding affinity to a second polypeptide. Those which bind selectively to the second polypeptide, such as a phosphatase, would be chosen for use in methods requiring a distinction between phosphatases or phosphatases and other polypeptides. Such methods could include, but should not be limited to, the analysis of altered phosphatase expression in tissue containing other polypeptides and assay systems using whole cells.

A peptide of the present invention can be used to produce antibodies or hybridomas. One skilled in the art will recognize that if an antibody is desired, such a peptide would be generated as described herein and used as an immunogen. The antibodies of the present invention include monoclonal and polyclonal antibodies, as well fragments of these antibodies, and humanized forms. Humanized forms of the antibodies of the present invention may be generated using one of the procedures known in the art such as chimerization or CDR grafting. The present invention also relates to a hybridoma which produces the

above-described monoclonal antibody, or binding fragment thereof. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

In general, techniques for preparing monoclonal  
5 antibodies and hybridomas are well known in the art  
(Campbell, "Monoclonal Antibody Technology: Laboratory  
Techniques in Biochemistry and Molecular Biology," Elsevier  
Science Publishers, Amsterdam, The Netherlands, 1984; St.  
Groth et al., J. Immunol. Methods 35:1-21, 1980). Any  
10 animal (mouse, rabbit, and the like) which is known to  
produce antibodies can be immunized with the selected  
polypeptide. Methods for immunization are well known in the  
art. Such methods include subcutaneous or intraperitoneal  
injection of the polypeptide. One skilled in the art will  
15 recognize that the amount of polypeptide used for  
immunization will vary based on the animal which is  
immunized, the antigenicity of the polypeptide and the site  
of injection.

The polypeptide may be modified or administered in an  
20 adjuvant in order to increase the peptide antigenicity.  
Methods of increasing the antigenicity of a polypeptide are  
well known in the art. Such procedures include coupling the  
antigen with a heterologous protein (such as globulin or  $\alpha$ -  
galactosidase) or through the inclusion of an adjuvant  
25 during immunization.

For monoclonal antibodies, spleen cells from the  
immunized animals are removed, fused with myeloma cells,  
such as SP2/0-Ag14 myeloma cells, and allowed to become  
monoclonal antibody producing hybridoma cells. Any one of a  
30 number of methods well known in the art can be used to

identify the hybridoma cell which produces an antibody with the desired characteristics. These include screening the hybridomas with an ELISA assay, western blot analysis, or radioimmunoassay (Lutz, et al., Exp. Cell Res. 175:109-124, 1988). Hybridomas secreting the desired antibodies are cloned and the class and subclass is determined using procedures known in the art (Campbell, "Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology", supra, 1984).

For polyclonal antibodies, antibody-containing antisera is isolated from the immunized animal and is screened for the presence of antibodies with the desired specificity using one of the above-described procedures. The above-described antibodies may be detectably labeled. Antibodies can be detectably labeled through the use of radioisotopes, affinity labels (such as biotin, avidin, and the like), enzymatic labels (such as horse radish peroxidase, alkaline phosphatase, and the like) fluorescent labels (such as FITC or rhodamine, and the like), paramagnetic atoms, and the like. Procedures for accomplishing such labeling are well-known in the art, for example, see (Stemberger, et al., J. Histochem. Cytochem. 18:315, 1970; Bayer, et al., Meth. Enzym. 62:308, 1979; Engval, et al., Immunot. 109:129, 1972; Goding, J. Immunol. Meth. 13:215, 1976). The labeled antibodies of the present invention can be used for in vitro, in vivo, and in situ assays to identify cells or tissues which express a specific peptide.

The above-described antibodies may also be immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such

as agarose and sepharose, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir et al., "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10, 1986; Jacoby et al., Meth. Enzym. 34, Academic Press, N.Y., 1974). The immobilized antibodies of the present invention can be used for in vitro, in vivo, and in situ assays as well as in immunochromatography.

Furthermore, one skilled in the art can readily adapt currently available procedures, as well as the techniques, methods and kits disclosed above with regard to antibodies, to generate peptides capable of binding to a specific peptide sequence in order to generate rationally designed anti-peptide peptides, for example see Hurby et al., "Application of Synthetic Peptides: Antisense Peptides", In Synthetic Peptides, A User's Guide, W.H. Freeman, NY, pp. 289-307 (1992), and Kaspczak et al., Biochemistry 28:9230-8 (1989).

#### **VII. An Antibody Based Method And Kit For Detecting a Polypeptide.**

The present invention encompasses a method of detecting a polypeptide in a sample, comprising: (a) contacting the sample with an above-described antibody, under conditions such that immunocomplexes form, and (b) detecting the presence of said antibody bound to the polypeptide. In detail, the methods comprise incubating a test sample with one or more of the antibodies of the present invention and assaying whether the antibody binds to the test sample.

Altered levels, either an increase or decrease, of a polypeptide in a sample as compared to normal levels may indicate disease.

Conditions for incubating an antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the antibody used in the assay. One skilled in the art will recognize that any one of the commonly available immunological assay formats (such as radioimmunoassays, enzyme-linked immunosorbent assays, diffusion based Ouchterlony, or rocket immunofluorescent assays) can readily be adapted to employ the antibodies of the present invention. Examples of such assays can be found in Chard, "An Introduction to Radioimmunoassay and Related Techniques" Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock et al., "Techniques in Immunocytochemistry," Academic Press, Orlando, FL Vol. 1(1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, "Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology," Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

The immunological assay test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as blood, serum, plasma, or urine. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted to the present invention.



A kit contains all the necessary reagents to carry out the previously described methods of detection. The kit may comprise: (i) a first container containing an above-described antibody, and (ii) second container containing a conjugate comprising a binding partner of the antibody and a label. In another preferred embodiment, the kit further comprises one or more other containers comprising one or more of the following: wash reagents and reagents capable of detecting the presence of bound antibodies.

Examples of detection reagents include, but are not limited to, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the chromophoric, enzymatic, or antibody binding reagents which are capable of reacting with the labeled antibody. The compartmentalized kit may be as described above for nucleic acid probe kits. One skilled in the art will readily recognize that the antibodies described in the present invention can readily be incorporated into one of the established kit formats which are well known in the art.

#### **VIII. Isolation of Natural Binding Partners of Polypeptides.**

The present invention also relates to methods of detecting natural binding partners capable of binding to a polypeptide. A natural binding partner of a polypeptide may be, for example, a substrate protein which is dephosphorylated as part of a signaling cascade. The binding partner(s) may be present within a complex mixture, for example, serum, body fluids, or cell extracts.

In general methods for identifying natural binding partners comprise incubating a substance with a phosphatase and detecting the presence of a substance bound to the phosphatase. Preferred methods include the two-hybrid  
5 system of Fields and Song (supra) and co-immunoprecipitation.

#### **IX. Identification of and Uses for Substances Capable of Modulating Polypeptide Activity**

10 The present invention also relates to a method of detecting a substance capable of modulating polypeptide activity. Such substances can either enhance activity (agonists) or inhibit activity (antagonists). Agonists and antagonists can be peptides, antibodies, products from  
15 natural sources such as fungal or plant extracts or small molecular weight organic compounds. In general, small molecular weight organic compounds are preferred. Examples of classes of compounds that can be tested for phosphatase modulating activity are, for example but not limited to,  
20 non-peptidyl compounds disclosed in Taylor, S. et al. (1998) Bioorganic and Medicinal Chemistry, 6:1457-1468, which is incorporated by reference, herein, compounds disclosed in Burke, Jr. et al. (1997) Current Pharmaceutical Design  
3:291-304 and Burke, Jr. et al. (1998) Biopolymers, 47:225-  
25 241, which are hereby incorporated by reference, herein, including any drawings.

In general, the method comprises contacting at least one polypeptide of the present invention with a test substance, measuring an activity of the polypeptide and  
30 determining whether the test substance modulates the

activity of the polypeptide. A change in activity may be manifested by increased or decreased phosphorylation of a phosphatase polypeptide or increased or decreased phosphorylation of a phosphatase substrate. The substance thus identified would produce a change in activity indicative of the agonist or antagonist nature of the substance.

The method also comprises incubating cells that produce phosphatases in the presence of a test substance and detecting changes in the level of phosphatase activity or phosphatase binding partner activity. A change in activity may be manifested by increased or decreased phosphorylation of a phosphatase polypeptide, increased or decreased phosphorylation of a phosphatase substrate, or increased or decreased biological response in cells. Biological responses can include, for example, proliferation, differentiation, survival, or motility. The substance thus identified would produce a change in activity indicative of the agonist or antagonist nature of the substance. Once the substance is identified it can be isolated using techniques well known in the art, if not already available in a purified form.

#### **X. Method for Treating a Disease or Disorder**

The present invention also relates to a method for treating a disease or disorder comprising the step of administering to a patient in need of such a treatment a substance that modulates phosphatases of the present invention.

Toxicity and therapeutic efficacy of substances, or compounds, can be determined by standard pharmaceutical procedures in cell cultures or experimental animals. The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD50/ED50. Compounds which exhibit large therapeutic indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized.

For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. For example, a dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the IC50 as determined in cell culture ( e.g., the concentration of the test compound which achieves a half-maximal disruption of the protein complex, or a half-maximal inhibition of the cellular level and/or activity of a complex component). Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by HPLC.

The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. (See e.g. Fingl et al., 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1 pl).

It should be noted that the attending physician would know how to and when to terminate, interrupt, or adjust administration due to toxicity, or to organ dysfunctions. Conversely, the attending physician would also know to  
5 adjust treatment to higher levels if the clinical response were not adequate (precluding toxicity). The magnitude of an administered dose in the management of the oncogenic disorder of interest will vary with the severity of the condition to be treated and with the route of  
10 administration. The severity of the condition may, for example, be evaluated, in part, by standard prognostic evaluation methods. Further, the dose and perhaps dose frequency, will also vary according to the age, body weight, and response of the individual patient. A program  
15 comparable to that discussed above may be used in veterinary medicine.

Depending on the specific conditions being treated, such agents may be formulated and administered systemically or locally. Techniques for formulation and administration  
20 may be found in "Remington's Pharmaceutical Sciences," 1990, 18th ed., Mack Publishing Co., Easton, PA. Suitable routes may include oral, rectal, transdermal, vaginal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous,  
25 intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections, just to name a few.

For injection, the agents of the invention may be formulated in aqueous solutions, preferably in  
30 physiologically compatible buffers such as Hanks's solution,

Ringer's solution, or physiological saline buffer. For such transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

5        Use of pharmaceutically acceptable carriers to formulate the compounds herein disclosed for the practice of the invention into dosages suitable for systemic administration is within the scope of the invention. With proper choice of carrier and suitable manufacturing  
10       practice, the compositions of the present invention, in particular, those formulated as solutions, may be administered parenterally, such as by intravenous injection. The compounds can be formulated readily using pharmaceutically acceptable carriers well known in the art  
15       into dosages suitable for oral administration. Such carriers enable the compounds of the invention to be formulated as tablets, pills, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated.

20       Agents intended to be administered intracellularly may be administered using techniques well known to those of ordinary skill in the art. For example, such agents may be encapsulated into liposomes, then administered as described above. Liposomes are spherical lipid bilayers with aqueous  
25       interiors. All molecules present in an aqueous solution at the time of liposome formation are incorporated into the aqueous interior. The liposomal contents are both protected from the external microenvironment and, because liposomes fuse with cell membranes, are efficiently delivered into the  
30       cell cytoplasm. Additionally, due to their hydrophobicity,

small organic molecules may be directly administered intracellularly.

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose. Determination of the effective amounts is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. The preparations formulated for oral administration may be in the form of, for example, tablets, dragees, capsules, or solutions.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is itself known, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain

substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility  
5 of the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical preparations for oral use can be obtained by combining the active compounds with solid excipients, optionally grinding a resulting mixture, and  
10 processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch,  
15 wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl  
20 pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate.

Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl  
25 pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound  
30 doses.



Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added.

Methods of determining the dosages of compounds to be administered to a patient and modes of administering compounds to an organism are disclosed in U.S. Application Serial No. 08/702,282, filed August 23, 1996 and International patent publication number WO 96/22976, published August 1 1996, both of which are incorporated herein by reference in their entirety, including any drawings, figures or tables. Those skilled in the art will appreciate that such descriptions are applicable to the present invention and can be easily adapted to it.

The proper dosage depends on various factors such as the type of disease being treated, the particular composition being used and the size and physiological condition of the patient. Therapeutically effective doses for the compounds described herein can be estimated initially from cell culture and animal models. For example, a dose can be formulated in animal models to achieve a circulating concentration range that initially takes into account the  $IC_{50}$  as determined in cell culture assays. The animal model

data can be used to more accurately determine useful doses in humans.

Plasma half-life and biodistribution of the drug and metabolites in the plasma, tumors and major organs can also be determined to facilitate the selection of drugs most appropriate to inhibit a disorder. Such measurements can be carried out. For example, HPLC analysis can be performed on the plasma of animals treated with the drug and the location of radiolabeled compounds can be determined using detection methods such as X-ray, CAT scan and MRI. Compounds that show potent inhibitory activity in the screening assays, but have poor pharmacokinetic characteristics, can be optimized by altering the chemical structure and retesting. In this regard, compounds displaying good pharmacokinetic characteristics can be used as a model.

Toxicity studies can also be carried out by measuring the blood cell composition. For example, toxicity studies can be carried out in a suitable animal model as follows: 1) the compound is administered to mice (an untreated control mouse should also be used); 2) blood samples are periodically obtained via the tail vein from one mouse in each treatment group; and 3) the samples are analyzed for red and white blood cell counts, blood cell composition and the percent of lymphocytes versus polymorphonuclear cells. A comparison of results for each dosing regime with the controls indicates if toxicity is present.

At the termination of each toxicity study, further studies can be carried out by sacrificing the animals (preferably, in accordance with the American Veterinary Medical Association guidelines Report of the American

Veterinary Medical Assoc. Panel on Euthanasia, Journal of American Veterinary Medical Assoc., 202:229-249, 1993).

Representative animals from each treatment group can then be examined by gross necropsy for immediate evidence of metastasis, unusual illness or toxicity. Gross abnormalities in tissue are noted and tissues are examined histologically. Compounds causing a reduction in body weight or blood components are less preferred, as are compounds having an adverse effect on major organs. In general, the greater the adverse effect the less preferred the compound.

For the treatment of cancers the expected daily dose of a hydrophobic pharmaceutical agent is between 1 to 500 mg/day, preferably 1 to 250 mg/day, and most preferably 1 to 50 mg/day. Drugs can be delivered less frequently provided plasma levels of the active moiety are sufficient to maintain therapeutic effectiveness.

Plasma levels should reflect the potency of the drug. Generally, the more potent the compound the lower the plasma levels necessary to achieve efficacy.

Examples of classes of compounds or compounds that may have phosphatase modulating activity are, for example but not limited to, non-peptidyl compounds disclosed in Taylor, S. et al. (1998) Bioorganic and Medicinal Chemistry, 6:1457-1468, which is incorporated by reference, herein, compounds disclosed in Burke, Jr. et al. (1997) Current Pharmaceutical Design 3:291-304 and Burke, Jr. et al. (1998) Biopolymers, 47:225-241.

### **XI. Method for Detection of a Phosphatase in a Sample as a Diagnostic Tool for a Disease or Disorder**

The invention also relates to a method for detection of a phosphatase in a sample as a diagnostic tool for a disease or disorder. The method may be practiced by contacting the sample with a nucleic acid probe which hybridizes under hybridization assay conditions to a nucleic acid molecule which encodes a phosphatase of the invention and detecting the presence or amount of a probe:target region as an indication of the disease. The method may also be practiced by comparing a nucleic acid target region encoding a phosphatase of the invention in a sample, to a control region and detecting differences in sequence or amount between the target region and the control region as an indication of the disease or disorder. The disease or disorder may be cancer, pathophysiological hypoxia such as seen in cardiac disfunction and vascular disorders including atherosclerosis, stenosis and stroke, myopathies, congenital muscle disorders, Papillon-Lefevre syndrome, Cowden disease, ectodermal dysplasia, Moebius syndrome, Bjornstad syndrome, Bannayan-Zonana syndrome, glioblastoma, schizophrenia and hamartomas. The cancer may be breast cancer, glioblastoma, urogenital cancer, , prostate cancer, head and neck cancer, lung cancer, synovial sarcomas, renal cell carcinoma, non-small cell lung cancer, hepatocellular carcinoma, pancreatic endocrine tumors, stomach cancer, colorectal cancer and thyroid cancer.

### **XII. Transgenic Animals**

Also contemplated by the invention are transgenic animals useful for the study of phosphatases activity in complex in vivo systems. A variety of methods are available for the production of transgenic animals associated with this invention. DNA sequences encoding phosphatases which can be injected into the pronucleus of a fertilized egg before fusion of the male and female pronuclei, or injected into the nucleus of an embryonic cell (e.g., the nucleus of a two-cell embryo) following the initiation of cell division (Brinster, et al., Proc. Nat. Acad. Sci. USA 82: 4438, 1985). Embryos can be infected with viruses, especially retroviruses, modified to carry inorganic-ion receptor nucleotide sequences of the invention.

Pluripotent stem cells derived from the inner cell mass of the embryo and stabilized in culture can be manipulated in culture to incorporate nucleotide sequences of the invention. A transgenic animal can be produced from such cells through implantation into a blastocyst that is implanted into a foster mother and allowed to come to term. Animals suitable for transgenic experiments can be obtained from standard commercial sources such as Charles River (Wilmington, MA), Taconic (Germantown, NY), Harlan Sprague Dawley (Indianapolis, IN), etc.

The procedures for manipulation of the rodent embryo and for microinjection of DNA into the pronucleus of the zygote are well known to those of ordinary skill in the art (Hogan, et al., supra). Microinjection procedures for fish, amphibian eggs and birds are detailed in Houdebine and Chourrout, *Experientia* 47: 897-905, 1991). Other procedures for introduction of DNA into tissues of animals are

described in U.S. Patent No., 4,945,050 (Sandford et al., July 30, 1990).

By way of example only, to prepare a transgenic mouse, female mice are induced to superovulate. After being  
5 allowed to mate, the females are sacrificed by CO<sub>2</sub> asphyxiation or cervical dislocation and embryos are recovered from excised oviducts. Surrounding cumulus cells are removed. Pronuclear embryos are then washed and stored until the time of injection. Randomly cycling adult female  
10 mice are paired with vasectomized males. Recipient females are mated at the same time as donor females. Embryos then are transferred surgically. The procedure for generating transgenic rats is similar to that of mice. See Hammer, et al., Cell 63:1099-1112, 1990).

15 Methods for the culturing of embryonic stem (ES) cells and the subsequent production of transgenic animals by the introduction of DNA into ES cells using methods such as electroporation, calcium phosphate/DNA precipitation and direct injection also are well known to those of ordinary  
20 skill in the art. See, for example, Teratocarcinomas and Embryonic Stem Cells, A Practical Approach, E.J. Robertson, ed., IRL Press, 1987).

In cases involving random gene integration, a clone containing the sequence(s) of the invention is co-  
25 transfected with a gene encoding resistance. Alternatively, the gene encoding neomycin resistance is physically linked to the sequence(s) of the invention. Transfection and isolation of desired clones are carried out by any one of several methods well known to those of ordinary skill in the  
30 art (E.J. Robertson, supra).

DNA molecules introduced into ES cells can also be integrated into the chromosome through the process of homologous recombination. Capecchi, Science 244: 1288-1292 (1989). Methods for positive selection of the recombination event ( e.g., neo resistance) and dual positive-negative selection ( e.g., neo resistance and gancyclovir resistance) and the subsequent identification of the desired clones by PCR have been described by Capecchi, supra and Joyner et al., Nature 338: 153-156, 1989), the teachings of which are incorporated herein. The final phase of the procedure is to inject targeted ES cells into blastocysts and to transfer the blastocysts into pseudopregnant females. The resulting chimeric animals are bred and the offspring are analyzed by Southern blotting to identify individuals that carry the transgene. Procedures for the production of non-rodent mammals and other animals have been discussed by others. See Houdebine and Chourrout, supra; Pursel, et al., Science 244:1281-1288, 1989); and Simms, et al., Bio/Technology 6:179-183, 1988).

Thus, the invention provides transgenic, nonhuman mammals containing a transgene encoding a phosphatase polypeptide or a gene effecting the expression of a phosphatase polypeptide. Such transgenic nonhuman mammals are particularly useful as an in vivo test system for studying the effects of introducing a phosphatase polypeptide, regulating the expression of a phosphatase polypeptide (e.g., through the introduction of additional genes, antisense nucleic acids, or ribozymes).

A "transgenic animal" is an animal having cells that contain DNA which has been artificially inserted into a

cell, which DNA becomes part of the genome of the animal which develops from that cell. Preferred transgenic animals are primates, mice, rats, cows, pigs, horses, goats, sheep, dogs and cats. The transgenic DNA may encode for a human  
5 phosphatase polypeptide. Native expression in an animal may be reduced by providing an amount of anti-sense RNA or DNA effective to reduce expression of the receptor.

### **XIII. Gene Therapy**

10 A phosphatase or its genetic sequences, both mutated and non-mutated, will also be useful in gene therapy (reviewed in Miller, Nature 357:455-460, (1992). Miller states that advances have resulted in practical approaches to human gene therapy that have demonstrated positive  
15 initial results. The basic science of gene therapy is described in Mulligan, Science 260:926-931, (1993).

In one preferred embodiment, an expression vector containing a phosphatase coding sequence or a phosphatase mutant coding sequence as described above is inserted into  
20 cells, the cells are grown in vitro and then infused in large numbers into patients. In another preferred embodiment, a DNA segment containing a promoter of choice (for example a strong promoter) is transferred into cells containing an endogenous gene sequence in such a manner that  
25 the promoter segment enhances expression of the endogenous phosphatase gene (for example, the promoter segment is transferred to the cell such that it becomes directly linked to the endogenous phosphatase gene).

The gene therapy may involve the use of an adenovirus  
30 containing phosphatase cDNA targeted to an appropriate cell



type, systemic phosphatase increase by implantation of engineered cells, injection with a phosphatase virus, or injection of naked phosphatase DNA into appropriate cells or tissues, for example neurons.

5        Expression vectors derived from viruses such as retroviruses, vaccinia virus, adenovirus, adeno-associated virus, herpes viruses, several RNA viruses, or bovine papilloma virus, may be used for delivery of nucleotide sequences (e.g., cDNA) encoding a recombinant phosphatase  
10 protein into the targeted cell population (e.g., tumor cells or neurons). Methods which are well known to those skilled in the art can be used to construct recombinant viral vectors containing coding sequences. See, for example, the techniques described in Maniatis et al.,  
15 Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, N.Y. (1989), and in Ausubel et al., Current Protocols in Molecular Biology, Greene Publishing Associates and Wiley Interscience, N.Y. (1989). Alternatively, recombinant nucleic acid molecules encoding protein  
20 sequences can be used as naked DNA or in a reconstituted system, e.g., liposomes or other lipid systems for delivery to target cells (See e.g., Felgner et al., Nature 337:387-8, 1989). Several other methods for the direct transfer of plasmid DNA into cells exist for use in human gene therapy  
25 and involve targeting the DNA to receptors on cells by complexing the plasmid DNA to proteins. See, Miller, supra.

In its simplest form, gene transfer can be performed by simply injecting minute amounts of DNA into the nucleus of a cell, through a process of microinjection. (Capecchi MR,  
30 Cell 22:479-88, 1980). Once recombinant genes are

introduced into a cell, they can be recognized by the cell's normal mechanisms for transcription and translation, and a gene product will be expressed. Other methods have also been attempted for introducing DNA into larger numbers of cells. These methods include: transfection, wherein DNA is precipitated with  $\text{CaPO}_4$  and taken into cells by pinocytosis (Chen C. and Okayama H, Mol. Cell Biol. 7:2745-52, 1987); electroporation, wherein cells are exposed to large voltage pulses to introduce holes into the membrane (Chu G., et al., Nucleic Acids Res., 15:1311-26, 1987); lipofection/liposome fusion, wherein DNA is packaged into lipophilic vesicles which fuse with a target cell (Felgner PL., et al., Proc. Natl. Acad. Sci. USA. 84:7413-7, 1987)); and particle bombardment using DNA bound to small projectiles (Yang NS. et al., Proc. Natl. Acad. Sci. 87:9568-72, 1990). Another method for introducing DNA into cells is to couple the DNA to chemically modified proteins.

It has also been shown that adenovirus proteins are capable of destabilizing endosomes and enhancing the uptake of DNA into cells. The admixture of adenovirus to solutions containing DNA complexes, or the binding of DNA to polylysine covalently attached to adenovirus using protein crosslinking agents substantially improves the uptake and expression of the recombinant gene. Curiel DT et al., Am. J. Respir. Cell. Mol. Biol., 6:247-52, 1992).

As used herein "gene transfer" means the process of introducing a foreign nucleic acid molecule into a cell. Gene transfer is commonly performed to enable the expression of a particular product encoded by the gene. The product may include a protein, polypeptide, anti-sense DNA or RNA,

or enzymatically active RNA. Gene transfer can be performed in cultured cells or by direct administration into animals. Generally gene transfer involves the process of nucleic acid contact with a target cell by non-specific or receptor mediated interactions, uptake of nucleic acid into the cell through the membrane or by endocytosis, and release of nucleic acid into the cytoplasm from the plasma membrane or endosome. Expression may require, in addition, movement of the nucleic acid into the nucleus of the cell and binding to appropriate nuclear factors for transcription.

As used herein "gene therapy" is a form of gene transfer and is included within the definition of gene transfer as used herein and specifically refers to gene transfer to express a therapeutic product from a cell in vivo or in vitro. Gene transfer can be performed ex vivo on cells which are then transplanted into a patient, or can be performed by direct administration of the nucleic acid or nucleic acid-protein complex into the patient.

In another preferred embodiment, a vector having nucleic acid sequences encoding a phosphatase is provided in which the nucleic acid sequence is expressed only in specific tissue. Methods of achieving tissue-specific gene expression as set forth in International Publication No. WO 93/09236, filed November 3, 1992 and published May 13, 1993.

In all of the preceding vectors, a further aspect of the invention is that the nucleic acid sequence contained in the vector may include additions, deletions or modifications to some or all of the sequence of the nucleic acid, as defined above.

In another preferred embodiment, a method of gene replacement is set forth. "Gene replacement" as used herein means supplying a nucleic acid sequence which is capable of being expressed in vivo in an animal and thereby providing  
5 or augmenting the function of an endogenous gene which is missing or defective in the animal.

The examples below are not limiting and are merely representative of various aspects and features of the present invention. The examples below demonstrate the  
10 isolation and characterization of the serine/threonine phosphatases of the invention.

**EXAMPLE 1: Isolation of cDNA clones Encoding Novel Mammalian**  
15 **Protein Phosphatases**

Identification and isolation of novel clones  
Novel protein tyrosine phosphatases (PTP) were identified from the public EST databases using a hidden Markov models (HMM; <http://pfam.wustl.edu/>) built mammalian and yeast  
20 phosphatase catalytic domain sequences. Dual specificity phosphatases were identified using an HMM model built from 93 DSPs from mammalian and non-mammalian sources (<http://pfam.wustl.edu/cgi-bin/getdesc?name=DSPc>). ESTs were translated in six open reading frames and were searched  
25 against the models. The public EST database was also searched by BLAST (Altschul, et al., (1997), Nucleic Acids Res. 25:3389-3402) with representative members of the various families, such as human DUS6, human MTM1, and human PTEN1. ESTs that had a score of at least 10 against the

HMM, or a E-value of less than 1.00 for the BLAST searches, were then masked for repetitive sequences and vectors and were clustered using MSA. The resulting contigs were searched against known phosphatases to identify EST clones that encode novel phosphatases. Full sequencing of EST and PCR fragments was carried out using a cycle sequencing Big-dye kit with AmpliTaq DNA Polymerase, FS (ABI, Foster City, CA). Sequencing reaction products were run on an ABI Prism 377 DNA Sequencer.

## 10 RESULTS

The following abbreviations were used for phosphatases:

DsPTP	Dual specificity protein phosphatase
DUS	Dual specificity phosphatase
GAK	cyclin G associated kinase
MKP	MAP Kinase phosphatase
MTM	Myotubular myopathy (myotubularin) phosphatase
PTEN	Phosphatase and tensin homolog

The following abbreviations were used for species

15

AT	Arabidopsis thaliana
CE	Caenorhabditis elegans
CI	Ciona intestinalis
DM	Drosophila melanogaster
H	Human
M	Murine
NT	Nicotiana tabacum
R	Rat
SC	Saccharomyces cerevisiae

SP Schizosaccharomyces pombe

Figure 3 discloses amino acid sequence alignments performed with the predicted ORF for the novel protein phosphatases against the non-redundant protein database (NRP database) as well as a database built with the novel phosphatases presented in this filing (Repository or "R"). Alignments were performed using the Smith-Waterman algorithm with a PAM 100 matrix table and gap open and extension penalties of 14 and 1, respectively. Figure 3 discloses "% Identity", "length of match", "Dbase" and "Hit sp", "Dbase hit" and "Hit Acc" refers to the calculated percent identity of each query against the best hits, along with the database source, species, description, and accession number of the match.

**EXAMPLE 2. Chromosomal Localization of Novel Mammalian Protein Phosphatases**

The chromosomal locations (CHR localization) for 8 of the 20 novel protein phosphatases are shown in Figure 1.

Several sources were used to find information about the chromosomal localization of each of the genes described in this patent. First, the accession number for the nucleic acid sequence was used to query the Unigene database. The site containing the Unigene search engine is:  
<http://www.ncbi.nlm.nih.gov/UniGene/Hs.Home.html>.  
Information on map position within the Unigene database is imported from several sources, including the Online Mendelian Inheritance in Man (OMIM,

<http://www.ncbi.nlm.nih.gov/Omim/searchomim.html>), The Genome Database

(<http://gdb.infobiogen.fr/gdb/simpleSearch.html>), and the Whitehead Institute human physical map

5 ([http://carbon.wi.mit.edu:8000/cgi-bin/contig/sts\\_info?database=release](http://carbon.wi.mit.edu:8000/cgi-bin/contig/sts_info?database=release)).

For example, searching Unigene with AA813123, an EST for a MKP-like phosphatase (SEQ ID#19), the following information is retrieved: X: DXS1061-DXS1039. The location of this gene  
10 on an "ideogram" of the cytogenetic map of chromosome X is also provided, showing that AA813123 maps to Xp11.4-q12. If Unigene has not mapped the EST, then the nucleic acid for the gene of interest is used as a query against databases, such as dbsts and htgs (described at

15 [http://www.ncbi.nlm.nih.gov/BLAST/blast\\_databases.html](http://www.ncbi.nlm.nih.gov/BLAST/blast_databases.html)) containing sequences that have been mapped already. The nucleic acid sequence is searched using BLAST-2 at NCBI (<http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-newblast>) and is used to query either dbsts or htgs. In addition to the  
20 Whitehead and GDB sites mentioned above, Stanford University maintains a useful site for chromosomal mapping from STS data

(<http://www-shgc.stanford.edu/RH/rhserverformnew.html>).

Matches in htgs are often resolved immediately because the  
25 genomic region hit is annotated in the htgs entry. If an exact match is found (defined roughly as 99% identity over a region of about 100 base pairs or longer, excluding any repetitive sequence), then the mapped position of the entry in the database is assigned to the original phosphatase  
30 query. Once a cytogenetic region has been identified by one

of these approaches, disease association is established by searching OMIM with the cytogenetic location. OMIM maintains a searchable catalog of cytogenetic map locations organized by disease. A thorough search of available literature for the cytogenetic region is also made using Medline (<http://www.ncbi.nlm.nih.gov/PubMed/medline.html>). References for association of the mapped sites with chromosomal abnormalities found in human cancer can be found in: Knuutila, et al., Am J Pathol, 1998, 152:1107-1123.

### Results of Chromosomal Mapping

#### **Seq ID#25 SGP033 maps to 2q33-q37.2**

This region has been associated with type I diabetes susceptibility. (Marron, et al. Diabetes. 2000 Mar; 49(3):492-9).

#### **Seq ID#17 LOC51207 (AA435513) maps to 10q21.3**

Allelic loss on chromosome 10q has been associated with human lung cancer tumor progression and metastatic phenotype (Petersen, et al., Br J Cancer. 1998;77(2):270-6) and with prostate tumor growth (Lacombe, et al., Int J Cancer 1996 Apr 22;69(2):110-3). In addition, Two tumor suppressive loci on chromosome 10 have been shown to be involved in human glioblastomas. (Genes Chromosomes Cancer 1995 Apr;12(4):255-61).

#### **Seq ID#29 MKP5 has been mapped to 1q32.1**

This region may be involved in renal collecting duct carcinoma (Steiner G, et al. Cancer Res. 1996 Nov



1;56(21):5044-6). This region has also been implicated in microcephaly and Van der Woude syndrome. (Kenwrick, et al, S, Hum Mol Genet. 1993 Sep;2(9):1461-2).

5 **Seq\_ID\_23 YVH1 (AA923158) maps to 1q21.2-q21.3**

Chromosomal aberrations of 1q21 have been linked to CNS disorders and cancer. Fananas et al. described a chromosomal fragile site at 1q21 in schizophrenic patients. (Am J Psychiatry (1997) 154:7-16). Zimonjic DB, et al.,  
10 described novel recurrent genetic imbalances in human hepatocellular carcinoma cell lines, identified by comparative genomic hybridization, mapping to this region (Hepatology (1999) 4:1208-14) Both loss and gain of distinct regions of chromosome 1q have been noted in primary  
15 breast cancer. (Bieche, et al., Clin Cancer Res. (1995) 1:123-7).

**Seq ID#19 AA813123 maps to Xp11.4-q12**

Translocations involving Xp11 have been associated with  
20 various human cancers. For example, most synovial sarcomas are characterized by a specific chromosomal translocation between Xp11 and 18q11.2. (Willeke, et al., Eur J Cancer 1998;34(13):2087-93). Perot et al. reported (Cancer Genet Cytogenet (1999) 110:54-6) two cases of papillary renal cell  
25 carcinoma (RCC) with a translocation between Xp11 and 1q21 in two female patients aged 9 and 29 years.

**Seq ID#21 AA915932 maps to 22q12.1-qter**

Frequent allelic deletions of 22q have been associated with  
30 human pancreatic endocrine tumors (Chung et al., Cancer Res

(1998) 58:3706-11, suggesting that the region contains one or more tumor suppressor loci.

**Seq ID#31 NP\_060746 maps to 11q12-q13.2**

- 5 This region has been implicated in adrenocortical carcinoma characterized by a high frequency of chromosomal gains and high-level amplifications. (Dohna M, et al., Genes Chromosomes Cancer. 2000 Jun 28(2):145-52).

10 **Seq ID#37 MTMR7 (AA663875) maps to 8p22**

- This region has been suggested to harbor a tumor suppressor gene involved in colon cancer (Lerebours, et al., Genes Chromosomes Cancer. (1999) 2:147-53). A genomic clone, AB020861, containing the gene encoding AA663875 was  
15 identified. AB020861 represents a human genomic DNA of 8p21.3-p22, and is annotated (Nakamura, Y. and Isomura, M., (in press) <http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=4003381&form=6&db=n&Dopt=g>) as  
20 containing an anti-oncogene of hepatocellular, colorectal, and non-small cell lung cancer.

**EXAMPLE 3. Expression analysis of Novel Mammalian Protein Phosphatases**

25 **GENE EXPRESSION ANALYSIS**

**Tissue Arrays**

"cDNA libraries" derived from a variety of sources were immobilized onto nylon membranes and probed with <sup>32</sup>P-labeled cDNA fragments derived from the gene(s) of interest. The  
5 sources of RNA are listed in Figure 3. They are: 1) Biochain Institute (Hayward, CA ; <http://www.biochain.com/main 3.html>); 2) Clontech (Palo Alto, CA, <http://www.clontech.com/>); 3) mammalian cell lines used by the National Cancer Institute (NCI) Developmental  
10 Therapeutics Program (<http://dtp.nci.nih.gov/>; can be ordered from ATCC: <http://www.atcc.org/catalogs.html>); 4) PathAssociates (<http://www.saic.com/company/subsidiaries/pai.html>; San Diego, California). The protocols for preparing cDNA arrays  
15 are detailed below.

**Preparation of total RNA from tissue and cultured cells.**

Stratagene RNA Isolation Kit (Stratagene #200345)

- 20 Homogenize 1.0g tissue in 10 ml solution D or resuspend tissue culture cell pellet from 10 100 mm dish<sup>(a)</sup> in 10ml solution D. Lyse cells by pipetting up and down. Add 1/10 volume 2 M sodium acetate (pH 4.0. Mix by inversion.
- 25 Add an equal volume phenol (pH 5.3-5.7) and mix by inversion. Add 1/5 volume chloroform-isoamyl alcohol and shake vigorously for 10 seconds. Incubate the tube on ice for 15 minutes.

Spin the tube at 10,000 x g for 20 minutes at 4°C. Transfer the upper aqueous phase to a fresh tube. Discard the lower phenol phase.

Add an equal volume of isopropanol to the aqueous phase and  
5 mix by inversion.

Incubate the tube for >1 hour at -20°C to precipitate the RNA.

Spin at 10,000 x g for 20 minutes at 4°C. After centrifugation, the pellet at the bottom of the tube  
10 contains the RNA.

Remove and discard the supernatant.

Dissolve the pellet in 3.0 ml of solution D.

Add 3.0 ml isopropanol and mix by inversion.

Incubate for 1 hour at -20°C.

15 Spin at 10,000 x g for 10 minutes at 4°C.

Remove and discard the supernatant.

Wash the pellet with 75% (v/v) ethanol ]DEPC-treated water (25%)].

Dry the pellet under vacuum for 2-5 minutes. Do not over  
20 dry.

Resuspend the RNA in desired volume DEPC-treated water.

Store at -80 C.

Depending on the amount of tissue or cells available, the protocol can be scaled up or down accordingly

25

**Preparation of polyA<sup>+</sup> mRNA from total human RNA.**

Reagents and Equipment

Oligotex mRNA Midi Kit (Qiagen).

Glycogen 2mg/mL in DEPC treated H<sub>2</sub>O

1. Thaw total RNA stored at  $-80^{\circ}\text{C}$ .
2. Mix 150  $\mu\text{L}$  RNA ( $\sim 0.3\text{--}1\text{ mg}$ ), 150  $\mu\text{L}$  2X binding buffer, and 55  $\mu\text{L}$  Qiagen Oligotex-dT resin.
- 5 3. Heat 3 minutes at  $65^{\circ}\text{C}$  to denature the RNA.
4. Cool to room temperature 10 minutes to allow annealing of mRNA to resin.
5. Pellet the resin containing bound mRNA by spinning for 2 minutes in a microfuge.
- 10 6. Resuspend the resin in 600  $\mu\text{L}$  of wash buffer by vortexing vigorously.
7. Pellet the resin by spinning for 2 minutes in a microfuge.<sup>(a)</sup>
8. Resuspend the resin in 600  $\mu\text{L}$  of wash buffer.
- 15 9. Transfer to a Qiagen spin column.
10. Spin out the wash buffer by centrifugation for 30 seconds in a microfuge.
11. Resuspend the resin in 33  $\mu\text{L}$  of  $80^{\circ}\text{C}$  elution buffer.<sup>(b)</sup>
12. Spin for 30 seconds and transfer the eluant containing  
20 the mRNA to a new tube.
13. Elute the remaining mRNA from the Oligotex resin in the spin column with two additional 33  $\mu\text{L}$  volumes of  $80^{\circ}\text{C}$  elution buffer.
14. Combine the three 33  $\mu\text{L}$  mRNA.
- 25 15. Add 1  $\mu\text{L}$  Glycogen, 10  $\mu\text{L}$  of 2.5 M sodium acetate and 220  $\mu\text{L}$  of 100% ethanol and place the tube at  $-80^{\circ}\text{C}$  overnight.
16. Pellet the mRNA by centrifugation for 30 minutes at  $4^{\circ}\text{C}$ .
- 30 17. Dry the mRNA pellet in a Speedvac.

18. Resuspend the pellet in 11  $\mu$ L of 1X TE (pH 8.0).<sup>(c)</sup>
19. Use 1.0  $\mu$ L for quantitation.<sup>(d)</sup>
20. Dilute to 0.1  $\mu$ g/ $\mu$ L.
21. Yield should be 20-25  $\mu$ g mRNA from 1.0 mg of total RNA.

5

- (a) Carry out the first wash in an eppendorf tube (e.g. by "batch wash") to remove particles and debris that clog the spin column (optional).
- (b) Heat the eppendorf tube containing the spin column to  
10 80°C for 30 seconds to assist in mRNA elution. Failure to do so will decrease mRNA yield!
- (c) Solutions are treated with 0.1% diethyl pyrocarbonate (DEPC) to inactivate ribonucleases.
- (d) Use 1.0  $\mu$ L of RNA in 0.1 mL Syber Green II ((Molecular  
15 Probes #S-7585) solution to determine concentration. The fluorescence is compared with the fluorescence of RNA standards to determine the concentration.

**Preparation of ssDNA from mRNA or total RNA.**

## Reagents and Equipment

Total RNA or mRNA

5 Primer CDS 57-mer (contains oligo dT 30-mer):

AAGCAGTGGTAACAACGCAGAGTACT<sub>30</sub>VN (V=A,G,C N=A,G,C,T)

Primer ML2G 30-mer:

AAGTGGCAACAGAGATAACGCGTACGCGGG

10 100 mM dATP, dCTP, dGTP, dTTP (Pharmacia #27-2050, -  
2060, -2070, -2080)SuperScript II RNase H<sup>-</sup> Reverse Transcriptase (Gibco  
BRL #18064-014)

1. In a 250  $\mu$ L PCR tube, mix 4.0  $\mu$ L of total mRNA <sup>(a)</sup> (100  
15 ng/ $\mu$ L) <sup>(b)</sup>, 1.0  $\mu$ L of oligo-dT containing primer CDS (10  
pm/ $\mu$ L).
2. Denature mRNA 2 minutes at 72°C and immediately place on  
ice for 2 minutes.
3. Spin briefly in a microcentrifuge at room temperature.
- 20 4. Add at room temperature; 2  $\mu$ L 5X First strand buffer  
(provided by BRL with the enzyme), 1 $\mu$ L DTT (20 mM), 1 $\mu$ L  
50X dNTP mix (10 mM each ) and 1  $\mu$ L SuperScript II  
reverse transcriptase (200 U/ $\mu$ L) for a total reaction  
volume of 10  $\mu$ L.
- 25 5. Mix by moving the pipette tip around gently.
6. Incubate the reaction at room temperature for 5 minutes.
7. Reverse transcribe the polyadenylated RNA for 1 hour at  
42°C in an air incubator.

8. Add 1  $\mu$ L of the ML2G oligo and mix with the pipette tip.
9. Incubate at room temperature for 5 minutes.
10. Incubate an additional 15 minutes at 42°C.
11. Terminate the reaction by adding 90  $\mu$ L of 10 mM Tris pH  
5 8.0 and freezing at -20°C for one hour or until needed  
for making double stranded cDNA.
- a. Total mRNA purified from total RNA using Oligotex-dT  
(Qiagen). As little as 10 ng of polyA<sup>+</sup> selected mRNA has  
been successfully used to make good quality ds DNA.
- 10 b. The same protocol is followed to make ssDNA from total  
RNA. To make ssDNA from total RNA it is preferable to  
use 1  $\mu$ g of total RNA, although good results can be  
obtained using total RNA amounts between 50 and 200 ng.
- c. If more than 200 ng of mRNA was used, add 440  $\mu$ L of Tris  
15 buffer.

Total RNA or mRNA was used as template in a reverse transcription reaction to generate single-stranded cDNAs (ss cDNA) that were tagged with specific sequences at each end.

20 An oligo dT primer containing a specific sequence (CDS: AAGCAGTGGTAACAACGCAGAGTACT<sub>30</sub>VN (V=A,G,C N=A,G,C,T)) anneals at the polyA track at the 3' end of the mRNA and the reverse transcriptase (MMLV RnaseH<sup>-</sup>) transcribes the antisense strand until it reaches the end of the RNA strand when it

25 adds additional C residues. If a primer (SMII: AAGCAGTGGTAACAACGCAGAGTACGCGGG or ML2G: AAGTGGCAACAGAGATAACGCGTACGCGGG) ending with 3 Gs is added, it anneals to the added Cs and the MMLV recognizes the rest of the primer sequence as template and continues

30 transcription. As a result, the synthesized cDNAs contain



specific sequence tags at both the 5' and the 3' end. When the 5' and the 3' ends are tagged with the same sequence (CDS and SMII) it is referred to as "symmetric". When the 5' end is tagged with a different sequence than the 3' end (CDS and ML2G) is referred to as "asymmetric". A double-stranded "cDNA library" is then generated by PCR amplification using the 3' PCR and ML2 primers (3' PCR: AAGCAGTGGTAACAACGCAGAGT and ML2: AAGTGGCAACAGAGATAACGCGT) that anneal to the added sequence tags.

10

#### Linear amplification of ds cDNA from cell lines and frozen tissues

##### Reagents and Equipment

15

Single stranded cDNA

Primer PCR 23-mer: AAGCAGTGGTAACAACGCAGAGT

Primer ML2 23-mer AAGTGGCAACAGAGATAACGCGT

100 mM dATP, dCTP, dGTP, dTTP (Pharmacia #27-2050, -2060, -2070, -2080)

20

Advantage 2 DNA polymerase (Clontech #8430-2)

Real time PCR (Roche, LightCycler or BioRad, iCycler)

Syber Green I (Molecular Probes #S-7585)

25 1 Single or double stranded cDNAs<sup>(a)</sup> are linearly amplified by PCR in the presence of fluorescent nucleotides to obtain double stranded cDNA probes. The optimal cycles

needed during amplification should have been predetermined for each sample.

2. Single stranded cDNA is linearly amplified by PCR to obtain double stranded cDNA. The linearity of the amplification is monitored by fluorescence in a real time PCR machine.
  3. Per 50.0  $\mu$ L reaction add: 1.0  $\mu$ L ss cDNA template, 5.0  $\mu$ L 10X Advantage 2 PCR Buffer, 1.0  $\mu$ L primer PCR (10 pm/ $\mu$ L), 1.0  $\mu$ L primer ML2 (10pm/mL), 1.0  $\mu$ L dNTP (10 mM each), 1.0  $\mu$ L Syber Green I (1:1,000 dilution of stock)<sup>(a)</sup>, 1.0  $\mu$ L Advantage 2 Polymerase Mix and 39.0  $\mu$ L H<sub>2</sub>O. Mix thoroughly and place onto real time PCR machine.
  4. Amplify according to the following regimen: 95°C for 1 min, then 35 cycles, 95°C for 5 sec, 65°C for 5 sec, 68°C for 6 min
  5. For each sample determine the optimal number of cycles<sup>(b)</sup>.
  6. Repeat step 2 using 5.0  $\mu$ L of template, omitting the Syber Green dye and adjusting the H<sub>2</sub>O to 35  $\mu$ L. Amplify for 2 cycles less than the determined number from step 4.
- (a) Syber Green is light sensitive.
- (b) The linear up slope of the fluorescence is the linear range of amplification. The optimal cycle number is the highest number within the linear portion of the curve. It is better to determine this number rather conservatively.
- (c) This step generates almost unlimited amounts of cDNA that can be used in generating fluorescent probes which in some cases might be desirable. If starting material is not in limited quantities, this step can be omitted

and proceed in generating fluorescent probes directly from ss cDNA.

#### **Arraying and Probing**

5 The amplified "cDNA libraries" were manually arrayed onto nylon membranes with a 384 pin replicator. Protocols for probe generation and hybridizing conditions can be found in Molecular Cloning : A Laboratory Manual (3 Volume Set), by T. Maniatis, and in manufactures protocols. Briefly, the DNA  
10 was denatured by alkali treatment, neutralized and cross-linked by UV light. The arrays were pre-hybridized with Express Hyb (Clontech) and hybridized with <sup>32</sup>P labeled probes generated by random hexamer priming of cDNA fragments (Stratagene PrimeIt Kit; Strtagene Corp); corresponding to  
15 the genes of interest. After washing, the blots were exposed to phosphorimaging cassettes and the intensity of the signal was quantified. The intensity of the spots was quantified using AIS software, Version 4.0, Rev 1.1, used in the 'DEFAULT' mode. (Imaging Research, Inc, St Catherines, Ontario, CA). The amount of the DNA on the arrays was also  
20 quantified by treating non-denatured or denatured arrays with Syber Green I or Syber Green II (Molecular Probes, <http://www.probes.com/>), respectively (1:100,000 in 50 mM Tris, pH8.0) for 2 minutes. After washing with 50mM Tris, pH8.0, the fluorescent emission was detected with a  
25 phosphorimager (Molecular Dynamics, <http://www.mdyn.com/>) and quantified using AIS software. The amount of the arrayed DNA was used to normalize the hybridization signal and the corrected values are tabulated in Figure 3.

## Results

The results of the microarray expression analysis of the protein phosphatases presented in this application is shown in Figure 3. Data presentation from left to right is as follows: "sample", the name of the sample; "source", where the sample was obtained; "tag", sym or asym depending on whether symmetric or asymmetric probe was used (see below for definitions of symmetric vs asymmetric); "type", lists tissue type (abbreviations: heme- hemotopoietic; pro - prostate; OV - ovarian; end - endocrine; mel - melanoma; neuro - neurological; leu - leukemia; col - colon, MG - mammary gland; "comments", comments on tissue source; "Tumor sym", indicates that the tissue is derived from a tumor, "sym" refers to the fact that the 5' and 3' primers used to make the sample are the same; "Normal Sym", indicates normal tissue was used to make the sample, with symmetric primers as described above; "Tumor 1°", indicates that primary tumor tissue was used to make the cDNA; "Tumor cells", indicates that these cDNA samples were made from cultured tumor cells; "Normal", indicates that these samples are derived from normal tissue or cell lines; "p53" refers to the status, mutant or wild-type, of the p53 gene in the source samples. Normalized expression values are presented for each gene referred to by its SEQ\_ID# on the subsequent columns. Genes represented in Figure 3 are: Actin, SEQ\_ID\_11\_AA374753, SEQ\_ID\_21\_AA915932, SEQ\_ID\_27\_AI031656, SEQ\_ID\_31\_NP\_060746 (G77-8-14), SEQ\_ID\_33\_NP\_060232 (AA232384), SEQ\_ID\_37\_MTMR7 (AA663875), and SEQ\_ID\_39\_AA493915.

By way of example, cDNAs made from RNA samples of a variety of tissue sources were spotted onto nylon membranes

and hybridized with radio-labeled probes derived from the phosphatase genes of interest. Referring to Figure 3, phosphatase gene sequences used include: SEQ\_ID\_11\_AA374753, SEQ\_ID\_21\_AA915932, SEQ\_ID\_27\_AI031656, SEQ\_ID\_31\_NP\_060746  
5 G77-8-14, SEQ\_ID\_33\_NP\_060232 AA232384, SEQ\_ID\_37\_MTMR7, and AA663875, SEQ\_ID\_39\_AA493915. As discussed herein, samples from normal tissues, tumor tissues, various cell lines, and P53 wild type and mutant were used to make the expression array. The relative gene expression levels of  
10 the tested phosphatase genes in various tissue sources were quantitated by measuring Syber Green I staining of hybridized signals. The numerical readings recorded in the figure were normalized to the hybridization result from ds cDNA or undenatured probes, after subtracting the background  
15 counts.

Together with the information of corresponding nucleic acid and amino acid sequences provided herein, the relevant expression levels in Figure 3 constitutes expression profiles of the phosphatase genes of interest in various  
20 tissue sources. Such expression profile data guides application of the treatment regime according to the present invention. For example, referring to the entry "primary renal cell adenocarcinoma (NCI 786-0)" in Figure 3, the levels of expression of SEQ\_ID\_11\_AA374753,  
25 SEQ\_ID\_21\_AA915932, SEQ\_ID\_27\_AI031656, SEQ\_ID\_33\_NP\_060232 AA232384 and SEQ\_ID\_39\_AA493915 are zero. The level of expression of SEQ\_ID\_31\_NP\_060746 G77-8-14 (203) is marginal. However, the level of expression of  
SEQ\_ID\_37\_MTMR7 AA663875 is significantly higher (2107).  
30 Such horizontal comparison reveals that the phosphatase gene

encoded by SEQ\_ID\_37\_MTMR7 AA663875 is implicated in renal cancer. That is, manipulation of the function activities of this gene may affect the cancerous condition of renal cancer. SEQ\_ID\_37\_MTMR7 AA663875 encodes homo sapiens  
5 myotubularin related protein 7, a MTM-like phosphatase as shown in Figure 2. Therefore, a method of treating the cancer condition connected to renal cancer according to the present invention can be, for example, to administer to the patient an agent that is capable of modulating the  
10 activities of the phosphatase activity of homo sapiens myobubularin related protein 7. The expression analysis according to the preferred embodiment of this invention thus confers specificity and effectiveness to the method of treatment disclosed.

15 These data also find applicability in a diagnostic setting. Referring to the same example, the entry primary renal cell adenocarcinoma (NCI 786-0) in Figure 3, the level of expression of SEQ\_ID\_37\_MTMR7 AA663875 is significantly higher (2107) compared to all other tested phosphatase genes  
20 (SEQ\_ID\_11\_AA374753, SEQ\_ID\_21\_AA915932, SEQ\_ID\_33\_NP\_060232 AA232384, SEQ\_ID\_27\_AI031656, SEQ\_ID\_31\_NP\_060746 G77-8-14, and SEQ\_ID\_39\_AA493915). This comparison thus demonstrates that a fair level of expression of the phosphatase gene encoded by SEQ\_ID\_37\_MTMR7 AA663875 correlates with the  
25 renal cancer condition. This gene may therefore be used as a diagnostic marker of the renal cancer condition, with certain reliability level. It is recommended, in this connection, that diagnostic tests to be run based on multiple markers to validate the test result and to increase  
30 the confidence level of the diagnosis derived as such.

Again, Figure 2 reveals that SEQ\_ID\_37\_MTMR7 AA663875 encodes homo sapiens myobubularin related protein 7, a MTM-like phosphatase. A method of diagnosing the cancer condition connected to neuroblastoma according to the present invention is, therefore, to contact a test sample, which may be collected from a patient, with a nucleotide probe which is capable of hybridizing to the nucleic acid sequence which encodes homo sapiens myobubularin related protein 7; and then to detect the presence of the hybridized probe:target pairs and to quantify the level of such hybridization as an indication of the cancer condition connected to neuroblastoma. The expression analysis according to the preferred embodiment of this invention thus confers specificity and effectiveness to the diagnostic method disclosed.

The presently representative of preferred embodiments are exemplary and are not intended as limitations on the scope of the invention. Changes therein and other uses will occur to those skilled in the art which are encompassed within the spirit of the invention are defined by the scope of the claims.

It will be readily apparent to one skilled in the art that varying substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention.

All patents and publications mentioned in the specification are indicative of the levels of those skilled in the art to which the invention pertains. All such documents mentioned herein, as well other various citations,

are incorporated by reference, even when not explicitly mentioned on the face of the document.

The invention illustratively described herein suitably may be practiced in the absence of any element or elements, 5 limitation or limitations which is not specifically disclosed herein. Thus, for example, in each instance herein any of the terms "comprising", "consisting essentially of" and "consisting of" may be replaced with either of the other two terms. The terms and expressions 10 which have been employed are used as terms of description and not of limitation, and there is no intention that in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are 15 possible within the scope of the invention claimed.

In particular, although some formulations described herein have been identified by the excipients added to the formulations, the invention is meant to also cover the final formulation formed by the combination of these excipients. 20 Specifically, the invention includes formulations in which one to all of the added excipients undergo a reaction during formulation and are no longer present in the final formulation, or are present in modified forms.

In addition, where features or aspects of the invention 25 are described in terms of Markush groups, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group. For example, if X is described as selected from the group consisting of bromine,



chlorine, and iodine, claims for X being bromine and claims for X being bromine and chlorine are fully described.

Other embodiments are within the following claims.

CLAIMS

What is claimed is:

1. An isolated, enriched or purified nucleic acid  
5 molecule encoding a polypeptide, wherein said nucleic acid molecule comprises a nucleotide sequence that
  - (a) encodes a polypeptide having the amino acid sequence set forth in SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 42, SEQ ID NO: 38 or SEQ ID NO: 40;
  - (b) is the complement of the nucleotide sequence of  
15 (a);
  - (c) hybridizes under highly stringent conditions to the molecule of (b) and encodes a naturally occurring polypeptide;
  - (d) encodes a polypeptide having the full length amino  
20 acid sequence set forth in SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 38, SEQ ID NO: 40 or SEQ ID NO: 42, except  
25 that it lacks one or more, but not all, of the amino acid numbers as set forth by the respective domain delimitations in any of the Figures;
  - (e) is the complement of the nucleotide sequence of (d);

(f) encodes a polypeptide having the amino acid sequence set forth in at least one of the respective sets of numbered amino acid residues set forth in any Figure;

5 (g) is the complement of the nucleotide sequence of (f);

(h) encodes a polypeptide having the full length amino acid sequence set forth in SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 38, SEQ ID NO: 40 or SEQ ID NO: 42, except that it lacks one or more, but not all, of the domains selected from the group consisting of an N-terminal domain, a phosphatase domain and a C-terminal domain;

15 (i) is the complement of the nucleotide sequence of (h);

(j) has the nucleotide sequence set forth in SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 23, SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 41, SEQ ID NO: 37 or SEQ ID NO: 39; or

(k) is the complement of the nucleotide sequence set forth in (j).

25

2. The nucleic acid according to Claim 1, further comprising a vector or promoter effective to initiate transcription in a host cell.

3. The nucleic acid molecule according to Claim 1,  
wherein said nucleic acid molecule is isolated, enriched or  
purified from a mammal.

5 4. The nucleic acid molecule according to Claim 3,  
wherein said mammal is a human.

5. A recombinant cell comprising a nucleic acid  
molecule, wherein said nucleic acid molecule encodes a  
10 polypeptide having the amino acid sequence set forth in at  
least one of the respective sets of numbered amino acid  
residues set forth in any Figure.

6. An isolated, enriched or purified polypeptide,  
15 wherein said polypeptide comprises an amino acid sequence  
having

(a) the amino acid sequence set forth in SEQ ID NO: 2,  
SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ  
ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ  
20 ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ  
ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ  
ID NO: 42, SEQ ID NO: 38 or SEQ ID NO: 40;

(b) the amino acid sequence set forth in SEQ ID NO: 2,  
SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 10, SEQ ID NO: 12,  
25 SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20,  
SEQ ID NO: 22, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30,  
SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 38, SEQ ID NO: 40 or  
SEQ ID NO: 42, except that it lacks one or more, but not all,  
of except that it lacks one or more, but not all, of the

amino acid numbers as set forth by the respective domain delimitations in any of the Figures

(c) the amino acid sequence set forth in at least one of the amino acid sequence set forth in at least one of the  
5 respective sets of numbered amino acid residues set forth in any Figure; or

(d) the amino acid sequence set forth in SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20,  
10 SEQ ID NO: 22, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32 or SEQ ID NO: 34, SEQ ID NO: 38, SEQ ID NO: 40 or SEQ ID NO: 42, except that it lacks at least one, but not all, of the following domains: an N-terminal domain, a C-terminal domain or a phosphatase domain.

15

7. The polypeptide according to Claim 6, wherein said polypeptide is isolated, purified or enriched from a mammal.

8. The polypeptide according to Claim 7, wherein said  
20 mammal is a human.

9. An antibody or antibody fragment having specific binding affinity to a polypeptide or a fragment thereof, wherein said polypeptide or fragment  
25 thereof has the amino acid sequence set forth in at least one of the respective sets of numbered amino acid residues set forth in any Figure.

10. A hybridoma which produces an antibody or antibody  
30 fragment according to Claim 9.

11. A method for identifying a substance that modulates the activity of a polypeptide, said method comprising the steps of

5 (a) contacting at least one polypeptide having the amino acid sequence set forth in at least one of the respective sets of numbered amino acid residues set forth in any Figure with a test substance;

(b) measuring an activity of the phosphatase; and

10 (c) determining whether the test substance modulates the activity of the phosphatase.

12. A method for identifying a substance that modulates phosphatase activity in a cell comprising the steps of

15 (a) expressing at least one phosphatase having the amino acid sequence set forth in at least one of the respective sets of numbered amino acid residues set forth in any Figure in a cell;

(b) adding a test substance to the cell; and

20 (c) monitoring

(i) a change in cell phenotype or

(ii) the interaction between the phosphatase and a natural binding partner.

25 13. A method for treating a disease or disorder comprising the step of administering to a patient in need of such a treatment a substance that modulates an activity of a phosphatase having the amino acid sequence set forth in at least one of the respective sets of numbered amino acid  
30 residues set forth in any Figure.

14. The method according to Claim 13, wherein said disease or disorder is selected from the group consisting of cancer, pathophysiological hypoxia, cardiac dysfunction and/or vascular disorders, myopathies, congenital muscle disorders, Papillon-Lefevre syndrome, Cowden disease, ectodermal dysplasia, Moebius syndrome, Bjornstad syndrome, Bannayan Zonana syndrome, schizophrenia and hamartomas.
15. The method according to Claim 14, wherein said cancer is selected from the group consisting of breast cancer, urogenital cancer, prostate cancer, head and neck cancer, lung cancer, synovial sarcomas, renal cell carcinoma, non-small cell lung cancer, hepatocellular carcinoma, pancreatic endocrine tumors, stomach cancer, glioblastoma, colorectal cancer and thyroid cancer.
16. The method according to Claim 15, wherein said substance modulates the activity of the phosphatase in vitro.
17. The method according to Claim 16, wherein said substance modulates the activity of the phosphatase by stimulating phosphatase activity.
18. A method for detection of a phosphatase in a sample as a diagnostic tool for a disease or disorder, wherein said method comprises the steps of
- (a) contacting said sample with a nucleic acid probe which hybridizes under hybridization assay conditions to a

nucleic acid which encodes a phosphatase having the amino acid sequence set forth in at least one of the respective sets of numbered amino acid residues set forth in any Figure;

- 5 (b) detecting the presence or amount of a probe:target region as an indication of the disease.

19. The method according to Claim 18, wherein said disease or disorder is selected from the group consisting of  
10 cancer, pathophysiological hypoxia, cardiac dysfunction and/or vascular disorders, myopathies, congenital muscle disorders, Papillon-Lefevre syndrome, Cowden disease, ectodermal dysplasia, Moebius syndrome, Bjornstad syndrome, Bannayan Zonana syndrome, schizophrenia and hamartomas.

15

20. The method according to Claim 19, wherein said cancer is selected from the group consisting of breast cancer, urogenital cancer, prostate cancer, head and neck cancer, lung cancer, synovial sarcomas, renal cell  
20 carcinoma, non-small cell lung cancer, hepatocellular carcinoma, pancreatic endocrine tumors, stomach cancer, glioblastoma, colorectal cancer and thyroid cancer.

21. A method for detection of a phosphatase in a sample  
25 as a diagnostic tool for a disease or disorder, wherein said method comprises the steps of

(a) comparing a nucleic acid target region of a nucleic acid, said nucleic acid encoding said phosphatase, in a sample to a control region, wherein said phosphatase has the  
30 amino acid sequence set forth in at least one of the



respective sets of numbered amino acid residues set forth in any Figure;

(b) detecting differences in sequence or amount between said target region and said control region as an indication of the disease or disorder.

22. The method according to Claim 21, wherein said disease or disorder is selected from the group consisting of cancer, pathophysiological hypoxia, cardiac dysfunction and/or vascular disorders, myopathies, congenital muscle disorders, Papillon-Lefevre syndrome, Cowden disease, ectodermal dysplasia, Moebius syndrome, Bjornstad syndrome, Bannayan Zonana syndrome, schizophrenia and hamartomas.

23. The method according to Claim 22, wherein said cancer is selected from the group consisting of breast cancer, urogenital cancer, prostate cancer, head and neck cancer, lung cancer, synovial sarcomas, renal cell carcinoma, non-small cell lung cancer, hepatocellular carcinoma, pancreatic endocrine tumors, stomach cancer, glioblastoma, colorectal cancer and thyroid cancer.

Ortho	Gene NAME	Sp	Id# no	Id# aa	FL Cat	Superfamily	Group	Family	Length NA	Length AA	ORF start	ORF end	ORF Length	DNA Repeats	SNP Position	CHR localization	Expression
1	AA023073	M	1	2	FL	Dual Phosphatase	DSP	CDC14	715	160	13	462	450	no	no	na	na
2	AA030322	M	3	4	FL	Dual Phosphatase	DSP	MKP	668	163	44	532	488	95-114	no	na	na
3	AA103395	M	5	6	FL	Dual Phosphatase	DSP	MKP	1067	184	3	593	591	no	no	na	na
4	AA114705	M	7	8	FL	Dual Phosphatase	DSP	MKP	697	186	1	567	597	no	no	na	na
5	AA274457	M	9	10	FL	Dual Phosphatase	DSP	MKP	628	220	1	560	660	321-338	no	na	na
6	AA374753	M	11	12	FL	Dual Phosphatase	DSP	MKP	1379	184	340	891	552	no	no	na	yes
7	AA398428	M	13	14	CAT	Dual Phosphatase	DSP	MKP	1020	219	1	654	654	454-472	no	na	na
8	AA422661	M	15	16	FL	Dual Phosphatase	DSP	MKP	1241	336	1	1017	1017	no	no	na	na
9	LOC51207	H	17	18	FL	Dual Phosphatase	DSP	MKP	904	199	68	671	594	no	376 (TAC); 395 (CTT); 537 (AAG)	10q21.3	na
10	AA513123	H	19	20	FL	Dual Phosphatase	DSP	MKP	808	180	154	723	570	11-28; 187-204	710 (TTC); 879 (TAA)	Xp11.4-q12	na
11	AA515932	H	21	22	FL	Dual Phosphatase	DSP	MKP	775	168	111	674	564	410-427	412 (TAA)	2q12.1-qter	yes
12	YVH1	H	23	24	FL	Dual Phosphatase	DSP	MKP	1251	340	12	1031	1020	no	115 (ATT)	1q21-q22	na
13	SGP033	H	25	26	FL	Dual Phosphatase	DSP	MKP	667	176	62	588	528	323-341; 541-559	no	2q33-q37.2	na
14	MKP-5	H	27	28	FL	Dual Phosphatase	DSP	MKP	901	217	66	736	661	no	650 (TGC); 875 (CTT)	1q32.1	yes
15	AI031656	H	29	30	FL	Dual Phosphatase	DSP	MKP	2050	482	285	1730	1446	102-162	502 (AAG)	1q12-q13.2	na
16	NP_060746	H	31	32	FL	Dual Phosphatase	DSP	MKP	1028	341	1	1023	1023	815-936	634 (CTG)	1q12-q13.2	yes
17	NP_060732	H	33	34	FL	Dual Phosphatase	DSP	MKP	3995	777	68	2396	2331	no	110 (TAC); 1098 (AG)	na	na
18	MTMR7	H	37	38	FL	Dual Phosphatase	DSP	MTM	2353	829	11	1897	1867	no	1688 (TAC)	9p22	yes
19	AA463915	H	39	40	CAT	Dual Phosphatase	DSP	PTEN	1300	400	1	1300	1300	no	368 (GAA); 375 (GAA); 371 (CTT)	na	yes
20	BAAG1172	H	41	42	FL	Dual Phosphatase	DSP	CDC14	694	150	35	484	480	no	478 (CTT)	na	na

FIGURE 1

Ortho	Gene Name	BP	DS no	DS as	Gene Family	Chr	Gene Start	Chr	Gene End	Rhodospirillum rubrum	Rhodospirillum rubrum	Length	ID match as	% Identity	% Similarity	Accession	Description
1	AA020371	M	3	4	DSF	10	148	X	148	X	X	150	172	85	98	NP_000293	Hypothetical protein FJ20442 (Homo sapiens)
2	AA030323	M	3	4	DSF	10	148	X	148	X	X	150	172	85	98	NP_000293	Hypothetical protein FJ20442 (Homo sapiens)
3	AA103585	M	5	6	DSF	43	141	X	141	X	X	143	165	76	92	NP_000497	SecY-1 like protein (Homo sapiens)
4	AA114705	M	7	8	DSF	43	141	X	141	X	X	143	165	76	92	NP_000497	SecY-1 like protein (Homo sapiens)
5	AA174457	M	8	10	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
6	AA174457	M	11	12	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
7	AA174457	M	13	14	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
8	AA174457	M	15	16	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
9	AA174457	M	17	18	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
10	AA174457	M	19	20	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
11	AA174457	M	21	22	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
12	AA174457	M	23	24	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
13	AA174457	M	25	26	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
14	AA174457	M	27	28	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
15	AA174457	M	29	30	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
16	AA174457	M	31	32	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
17	AA174457	M	33	34	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
18	AA174457	M	35	36	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
19	AA174457	M	37	38	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
20	AA174457	M	39	40	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
21	AA174457	M	41	42	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
22	AA174457	M	43	44	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
23	AA174457	M	45	46	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
24	AA174457	M	47	48	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
25	AA174457	M	49	50	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
26	AA174457	M	51	52	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
27	AA174457	M	53	54	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
28	AA174457	M	55	56	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
29	AA174457	M	57	58	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
30	AA174457	M	59	60	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
31	AA174457	M	61	62	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
32	AA174457	M	63	64	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
33	AA174457	M	65	66	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
34	AA174457	M	67	68	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
35	AA174457	M	69	70	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
36	AA174457	M	71	72	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
37	AA174457	M	73	74	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
38	AA174457	M	75	76	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
39	AA174457	M	77	78	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
40	AA174457	M	79	80	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
41	AA174457	M	81	82	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
42	AA174457	M	83	84	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
43	AA174457	M	85	86	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
44	AA174457	M	87	88	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
45	AA174457	M	89	90	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
46	AA174457	M	91	92	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
47	AA174457	M	93	94	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
48	AA174457	M	95	96	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
49	AA174457	M	97	98	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
50	AA174457	M	99	100	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
51	AA174457	M	101	102	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
52	AA174457	M	103	104	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
53	AA174457	M	105	106	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
54	AA174457	M	107	108	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
55	AA174457	M	109	110	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
56	AA174457	M	111	112	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
57	AA174457	M	113	114	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
58	AA174457	M	115	116	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
59	AA174457	M	117	118	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
60	AA174457	M	119	120	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
61	AA174457	M	121	122	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
62	AA174457	M	123	124	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
63	AA174457	M	125	126	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
64	AA174457	M	127	128	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
65	AA174457	M	129	130	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
66	AA174457	M	131	132	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
67	AA174457	M	133	134	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
68	AA174457	M	135	136	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
69	AA174457	M	137	138	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
70	AA174457	M	139	140	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
71	AA174457	M	141	142	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
72	AA174457	M	143	144	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
73	AA174457	M	145	146	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
74	AA174457	M	147	148	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
75	AA174457	M	149	150	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
76	AA174457	M	151	152	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
77	AA174457	M	153	154	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
78	AA174457	M	155	156	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
79	AA174457	M	157	158	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
80	AA174457	M	159	160	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
81	AA174457	M	161	162	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
82	AA174457	M	163	164	DSF	43	202	X	202	X	X	202	220	78	93	NP_000577	UAS11 (Homo sapiens)
83	AA174457	M	165	166	DSF	43	202	X									

Tumor-adj	Normal-adj	Tumor - 16	Tumor cells	Normal	p23	actb	img_11	img_21	img_27	img_29	img_30	img_37	img_39
218													
222						2,315,182	6,835	0	0	431	156	0	60
						2,484,020	0	78	775	330	0	2,818	1,978
2						801,610	0	0	0	0	486	770	0
332						3,940,861	1,069	0	0	0	0	0	0
1						1,802,771	0	250	62	250	796	1,798	3,693
277						1,803,125	0	243	210	0	849	0	0
3						2,131,351	0	660	0	0	0	0	0
279						4,141,910	3,138	0	4,774	1,051	343	11,214	7,209
6						2,052,387	0	30	82	0	218	568	78
342						2,052,387	0	30	82	0	218	568	78
7						1,575,769	1,069	0	3,065	1,773	114	1,115	8,42
344						3,332,648	15,160	646	8,744	1,040	3,645	35,840	33,338
9						3,332,648	15,160	646	8,744	1,040	3,645	35,840	33,338
42						2,884,827	4,005	0	8,187	10,009	1,834	2,720	18,319
210						2,238,319	3,602	820	3,187	10,009	1,834	2,720	18,319
238						3,221,184	2,318	0	1,022	1,022	1,022	1,022	1,022
11						3,221,184	2,318	0	1,022	1,022	1,022	1,022	1,022
48						3,221,184	2,318	0	1,022	1,022	1,022	1,022	1,022
13						1,458,515	0	229	431	362	0	6,534	224
61						1,458,515	0	229	431	362	0	6,534	224
219						1,722,624	201	102	0	4,182	862	1,117	10,603
16						1,683,323	464	102	134	0	4	0	0
53						1,683,323	464	102	134	0	4	0	0
38						3,000,225	0	0	1,861	481	189	7,377	2,059
18						1,688,038	177	128	0	185	656	283	185
37						1,211,623	3,203	0	2,069	142	1,598	20,351	6,040
21						2,418,844	878	0	348	431	1,108	2,118	1,328
287						3,676,622	17,842	844	7,868	3,259	1,261	1,887	318
23						712,448	3,275	0	615	486	331	15,716	2,889
38						7,438,238	2,975	0	618	331	802	885	28
38						2,911,688	0	0	0	25	0	869	9
61						2,008,110	8,710	0	8,778	117	0	8,211	10,877
303						8,820,235	3,781	0	301	218	88	22	8,473
35						1,352,624	0	0	4,089	300	862	12,538	10,185
37						3,026,020	1,028	0	0	21	118	0	0
307						3,888,984	3,328	0	489	252	80	878	18
10						1,832,543	0	216	2,141	723	2,810	18,581	5,848
212						1,871,764	0	0	0	188	0	1,823	608
264						4,838,510	830	0	170	0	1,446	12,348	18,802
300						2,184,151	821	0	182	143	0	85	0
187						2,188,005	1,251	813	2,133	468	1,311	0	38
353						8,817,864	0	568	8,215	341	0	8,166	653
353						1,711,488	0	0	0	80	231	33	0
252						2,378,388	4,270	0	2,423	228	0	11,874	4,168
20						6,084,717	3,118	83	413	828	4,172	2,116	21
294						1,184,228	23,835	711	3,118	810	0	5,264	2,337
22						1,118,748	1,818	181	3,108	810	0	10,384	2,360
236						4,261,288	0	0	78	405	78	1,248	18
214						825,658	1,812	74	4,188	23	0	8,360	3,847
23						3,188,146	478	86	798	0	1,485	2,028	846
44						1,728,020	0	0	212	328	30,891	38,168	886
189						5,848,910	17,646	0	3,788	10,660	23,827	3,247	0
183						8,820,595	25,344	460	8,360	4,839	14,827	18,827	200
40						4,128,014	2,823	55	481	480	0	3,123	3,156
228						3,128,873	4,823	224	8,115	788	0	1,274	1,178
314						8,840,486	34,310	0	8,845	1,215	0	7,178	47
316						8,817,864	0	0	1,866	765	0	8,388	6,732
318						8,021,352	11,877	268	0	402	1,483	788	1,641
319						3,172,780	447	0	3,878	883	22,341	4,408	0
32						8,411,550	18,818	778	8,681	1,288	2,887	1,321	754
182						4,316,828	3,324	243	3,037	582	0	24,115	744
182						10,338,816	29,021	478	10,860	1,472	1,148	3,880	568
182						2,248,184	0	0	1,171	1,171	0	2,107	0
182						8,881,882	8,881	0	8,784	1,420	370	0	0
182						4,782,482	8,418	82	1,246	818	0	0	0
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													
182													

**FIGURE 3 (1 OF 4)**

[illegible]

**FIGURE 3 (2 OF 4)**

[illegible]

**FIGURE 3 (3 OF 4)**

[illegible]

**FIGURE 3 (4 OF 4)**

>SeqID\_1\_AA023073\_m

ACTCCCCCACGATGGGCGTGCAACCCCCCAACTTCTCCTGGGTGCTTCCGGGACGGCTGG  
CCGGAAGTGGCGTTGCCCGGCTGCCCGCGCACTACCAAGTTCCTGCTGGACAGGGTGTGC  
GGACCTGGTGTCCCTGACGGAGCGCGGACCCCTCACAGTGACAGCTGTCCCGGCTCA  
CGCTGCACCGAATGCGCATCCCTGACTTTTGCCCGCGCTCCCGGAACAGATCGACCAATT  
TGTGAAGATCGTGGACGAGGCCAATGCCCGGGAGAGGCTGTTGGAGTGCAGTGTGCCCT  
AGGCTTTGGCCGCACTGGCACCATGCTAGCCTGCTACTTGGTGAAGGAGCGGGCTTTGGC  
CGCAGGAGATGCCATTGCTGAGATCCGCGCGCTGCGACCAAGGATCCATTGAGACGTATGA  
ACAGGAGAAGGCCGTCTTCCAGTTCTACCAGCGAACAATAAGGAGTCAACAAGCCCG  
CCTTTCCCTCCCAACTCCTGCGGCCAGGGAGGAAGGGAGTGAATAAAGTACTGCA  
TCCTTCAGGTCCCTCTGACTCTATTGGACAAAAGTAGTCTTCCCAAGCCATAACGTG  
GCCGCGAGGATGGCCGAGACCCACAAAATGAGGTAATAACTGATAAGAACTCATCAC  
CGCTGCATAGCATGTACACAGCACTCCCAATACATCTGGGTGGTTGAAAAGAC

>SeqID\_3\_AA030322\_m

CGCAACTCGCGGCCCGTAGATGGCCCTAGGGTCGCAGGGGTTATGGGCACCTCGGAGGC  
TGCACCGCGCCCTTTGCGCGCGCTCGCCCCGCGCTTTCATCGGGAATGCGCGAGCCCGG  
GGTGGCAGCGGAGCTGCTGGTGCAGCGCGGCATCACTTTGTGCGTCAATGTCTCCCGCCAG  
CAGCCCGGGCCGCGCGCGCCGAGTGGCGGAACACGCGTACCCGTGTTTCGACGACCCA  
GCTGAGGACCTGCTGACACACCTGGAGCCCACTGTGCGCCATGGAAGCCGCGGTGCGC  
GACGGCGGCTCCTGTCTCGTACTGCAAGAACGGCCGAGTCGCTCAGCCGCGGTCTGC  
ACCGCCTACCTAATGCGGCACCGCGGCCACAGCCTGGATCGCGCCTTCCAGATGGTGAAG  
AGCGCCCGCCCGGTAGCCGAGCCCAATTTGGGGTCTTGGGCTCAGCTGCAGAAAGTACGAG  
CAGACCTTCAGGCCCAGGCCATCCTGCCCGGGAGCCATTGATCCGGAGTAAGCCGAC  
TGTTCCGCTGCTGGGTGACCAAGCGTCTATACTGAAAGGAAGTGTCCCTTCCCTCCTTTT  
CTATTAGGCAGCTGGCTTTGGGTGTGCCCCATCTTGATGGTAGTACAGGAACGTCTACTG  
AGTAGGAGGACTTCGTTTATTCATCATGTTTGGACCAAAATCCAAACCAGCACGTTTATAGT  
AGAGAAATTGAGTGAAGGATAGTCTGGGAAGCCTACGAACGGTTGATAGCGAGTGATAG  
ATCAGAGTCTAGCTGCCTACTCCAAGGGAGTGCTGGGTTTGTAGGCAGAACCTATCTGT  
CTCCTGAACTTCTGGTCCCTTAGAAATGAACATAGAGTCTCCAGCAGGAGCTCATGGGCC  
CACCTTTGGCTTACAGCTGCTGTGCCATGGAAGGGAGGCCGTGCAGACTGCAGCTGAGCG  
ACT

>SeqID\_5\_AA103595\_m

CCCGCTCCGGCGAGCGCGGAGCGCGAGCCCGCGCGGCCATGGGGAGTGGGATGAGC  
CAGATCCTGCCGGGCTGTACATTGGCAACTTCAAAGACGCAAGAGATGCAGAACAGTTG  
AGCAGGAACAAGGTGACACACATTCTTCTGTGCAGGATACTGCCAGGCCCATGTTGGAG  
GGAGTTAAATACCTGTGTATTCCAGCGGCAGACACACCATCTCAAACCTGACAAGACAT  
TTCAAAGAAAGCATTAAATTCATTATGAGTGCCGACTCCAGGGTGAGAGCTGTCTGTAC  
ATTGCTGGCTGGGGTCTCCAGGAGTGTGACATTGGTGATCGCATACATCATGACTGTAC  
CGACTTTGGCTGGGAAGATGCCTTGACACTGTTCTGTGCGGGGAGGTCTGTGCCAACCCC  
AACCTGGGCTTTCAAAGGCAGCTGCAGGAGTTTGAGAAACATGAAGTGCACCAAGTATCGG  
CAATGGCTGAGAGAAGAGTATGGAGAGAACCCTTTGCGGGATGCAGAAGAAGCCAAAA  
TATTCTGGCTGCCCCGGGAATTCTGAAGTACTGGGCCTTTCTCAGAAGACTGTAATGTACC  
TGAAGTTTCTGAAATATTGCAAAGTTCAGGCTGGTGTGCTGCCAAAAAGAAAGTATGTAA  
AGTTTATTTTTAAGAATCCAATAGTGATTTGTATACTTGTTTTTTTTCAATTTAAACCAA  
TGCAATGATAATCATGTTGGAATATGTTAAGATCTATGGATATTCTGTAGCAAGAGAAAAT  
ATCTTTGCCCTTAACCTCACTGCTGTGGTTGTTCTTGGACCTGACCGATGCTCATACAATA  
TCTCAAGAGCCCTGTCTGTTTCGTAATAGTAATACTTCTCATGAACACTACCAAGGAGG  
AAGCCTGCACCTGGGAAGTGTGAGTGTGAGCTCTGCCCTCCTGTTAAGTTCTCCAGCTCT  
AGACATGTCTCTTGGTGTGTGTTTATCTACTGGTGTATTCTATATGGTAGAATTACCAA  
AGCTATTAGATTCTTAATAAAGGGCAAATCAACC

>SeqID\_7\_AA144705\_m

ATGGACTCGCTACAGAAGCAGGAACCTTCGGAGGCCAAAGATTTCATGGGGCAGTCCAGGTG  
TCCCCCTACCAAGCCACCCACACTGGCCTCTCTGCAGCGATTGCTGTGGGTCCGTCCGACTG  
CCCACTGACCCACATCAATGAGGTCTGGCCCAACCTTTCTTGGGAGATGCGTATGCTGC  
CAGAGACAAGGGTCTGCTAATCCAGCTGGGCATTACCCATGTTGTAATGTGGCTGCGGG  
CAAGTTCAGGTGGACACAGGTGCCAAGTTCTACCGTGGAACACCTCTGGAGTACTATGG  
CATTGAGGCTGATGACAACCCCTTCTTGACCTCAGCGTCCACTTTCTGCCTGTTGCTCGTT



ACATCAGAGATGCCCTCAATATCCCCGAAGCCGAGTGCTGGTCCACTGCGCTATGGGGG  
TGAGTCGCTCTGCCACAAATGTCTTGGCCTTCCTCATGATCTTCGAGAACATGACACTGGT  
AGATGCCATCCAGACGGTGCAGGCCACCGAGATATCTGTCCAACTCAGGCTTCCTCCG  
ACAGCTCCAGGTTCTGGACAACAGGCTGAGGCCGGAACAGGAAGACTCTGA

>SeqID\_9\_AA274457\_m

ATGCACTCCCTGAACCAAGAAATCAAAGCATTCTCCCGGGATAATCTCAGGAAGCAGTGC  
ACCAGGGTGACCACGCTAACTGGAAAGAACTTATAGAAACCTGGGAAGATGCCACAGTT  
CATGTTGTGGAGACAGAGCCCAGCGGTGGGGGTGGTTGTGGCTATGTGCAGGACCTTACC  
TTGGACCTGCAAGTTGGCGTTATTAAGCCCTGGTTGCTTCTGGGGTCACAGGATGCTGCTC  
ATGACCTGGAGCTACTGAGAAAGCATAAGGTGACTCATATTCTCAATGTTGCATATGGAG  
TTGAAATGCTTTCCTCAGTGAGTTTACATATAAGACCATTCTATACTGGATGTGCCTGA  
AACCAATATCCTGTCTTATTTTCCAGAATGTTTGGAGTTTATTGAGCAAGCAAAACTGAAG  
GATGGCGTGGTTCTCGTGCCTGTAAATGCAGGTGTTTCCAGGGCTGTGCAATTGTCTATTG  
GCTTCTCATGAGTTCTGAAGAAGCCACTTTCACCACTGCCCTGTCGTTGGTGAAAGAGGC  
GAGACCATCCATATGTCCGAATCCTGGCTTCATGGAACAACTCCGCACCTACCAAGTAGG  
CAAGGAGAGCAATGGAGGTGACAAAGTGCCCCGCGGAGGACACGACCGGTGGTCTGTGAT  
CTGTACTCCAGCAGAGGCAACGACTTCTGCATCAGACTCTGTCTCTTGCCTGTGTTTTG  
GAAGGAAACTTGGAAAACTTCCCTTTTCTGTGTTCTTTTACCAGTGGAATGAACGTTACC  
TTTGTGCGCCTGAATTAATACATTTTAAAGTTTACCTTTTC

>SeqID\_11\_AA374753\_h

GGCGGGCGGACGAGGAGGGACGCTGGGCCTGCCCGGTNGCGCACGGGGGCGGGGACCG  
GCAAGGCGGGACCAATTTCCCGGCATAGGCTCCGGTGCCCTGCCCGGCTCCCGCGGGAA  
GTTCTAGGCCGCGGCACAGAAAGCCCTGCCCTCCACGCCGGGTCTCTGGAGCGCCCTGGG  
TTGCCCGGCCGCTCCCTGCCGCTGACTTGTTGACACTGCGAGCACTCAGTCCCTCCCGCGC  
GCCTCCTCCCGCCCGCCCGCCGCTCCTCCTCCTGTAACATGCCATAGTGCGCCTGCGA  
CCACACGGCCGGGGCGCTAGCGTTCCGCTTCAGCCACCATGGGGAATGGGATGAACAAGA  
TCCTGCCCGGCCTGTACATCGGCAACTTCAAAGATGCCAGAGACGCGGAACAATTGAGCA  
AGAACAAGGTGACACATATTCTGTCTGTCCACGATAGTGCCAGGCCTATGTTGGAGGGAG  
TTAAATACCTGTGCATCCAGCAGCGGATTACCATCTCAAAACCTGACAAGACATTTCAA  
AGAAAGTATTAAATTCATTCACGAGTGCCCGGCTCCGCGGTGAGAGCTGCCTTGTACACTG  
CCTGGCCGGGGTCTCCAGGAGCGTGACACTGGTGATCGCATACATCATGACCGTCACTGA  
CTTTGGCTGGGAGGATGCCCTGCACACCGTGCGTGCTGGGAGATCCTGTGCCAACCCCAA  
CGTGGGCTTCCAGAGACAGCTCCAGGAGTTTGAGAAGCATGAGGTCCATCAGTATCGGCA  
GTGGCTGAAGGAAGAATATGGAGAGAGCCCTTTCAGGATGCAGAAGAAGCCAAAAACA  
TTCTGGCCGCTCCGGGAATCTGAAGTTCTGGGCCTTCTCAGAAGACTGTAATGTACCTG  
AAGTTTCTGAAATATTGCAAAACCCACAGAGTTTAGGCTGGTGCTGCCAAAAGAAAAGCA  
ACATAGAGTTTAAAGTATCCAGTAGTGATTTGTAAACTTGTTTTTCATTTGAAGCTGAATAT  
ATACGTAGTCATGTTTATGTTGAGAACTAAGGATATTCTTTAGCAAGAGAAAAATTTTCC  
CCTTATCCCCACTGCTGTGGAGGTTTCTGTACCTCGCTTGGATGCCTGTAAGGATCCCGGG  
AGCCTTGCCGCACTGCCTTGTGGGTGGCTTGGCGCTCGTGATTGCTTCTGTGAACGCCTC  
CCAAGGACGAGCCAGTGATGTTGTGTGGCGTGAACCTGCCCCGTGTGTTCTCAAATTTCC  
CAGCTTGGGAAATAGCCCTTGGTGTGGGTTTATCTCTGGTTTGTGTTCTCCGTGGTGAA  
TTGACCGAAAGCTCTATGTTTTCGTTAATAAAGGGCAACTTAGCCAAAGTTT

>SeqID\_13\_AA396428\_m

GAACCCCTGGTATTGAAAGGGGGACTCAGTAGTTTAAACAGAACCATGGAAACCTCTGT  
GACAACTCCCTCCAGCTCCAAGAGTGCCGGGAGGTGGGGGTGGTGATCTGCGGCCTCG  
AGCATGTACCTCAGTCTGTCCCCACCACCCCTGACATCGAGAACGCAGAGCTAACGCCC  
ATCCTGCCCTTCTGTTCTCCTCGGCAATGAGCAGGATGCTCAGGACCTAGACACCATGCAGA  
GGCTCAACATCGGCTATGTCAACGTCAACACGCACCTTCTCTGTACCATATGAGAA  
AGGCCTTCTCAACTACAAGAGGCTGCCAGCCACAGACAGCAACAAACAGAACCTGCGGC  
AGTACTTTGAAGAGGCCTTCGAGTTTCATCGAGGAAGCTCACCAGTGTTGGGAAGGGCCTTC  
TCATCCACTGCCAGGCCGCGTGTCCCGATCCGCCACCATCGTCATCGCCTACTTGATGAA  
GCACACACGGATGACCATGACTGACGCTTACAAATTCGTCAAAGGCAAAACGACCAATTAT  
TTCCCCGAACCTCAACTTCATGGGGCAGTTGCTGGAATTTGAGGATGACCTAAACAACGG  
CGTGACGCCAAGAATCCTTACACCAAAGCTCATGGGCATGGAGACAGTTGTGTGACAACG  
GGCAGGACGGAAGGGCTGTGCTCTCTCAGGAGACGAAAAGGAGGGAAGGTGGATTCTA  
GCTGTGCTCTTCTTTCCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT

FIGURE 4 (2 OF 8)

TTCTTTTCTTTCTCTTTCTTTCTCTTTTTTAAGGTTGGGAGTGTTTGTGAAAGTAAACA  
AACTTGTCTAAAGACTCTATTTTAAATAAGTGTAAGAAGACTGTAACTTTGTATGCGAATG  
AGACCCACTTCTCTTAAACTGGTATTGCAGTTAGGTTAAAGAGGTCCTTTTGTGTTT  
GTTTTTTAAAGCCAACCCATAAAAAATTATAAAACTTGGTCTTTCC

>SeqID\_15\_AA422661\_m

ATGGTTGGAAGCGCAGGGTACTAACCATGGCTGCGAGCGCCAGGCTCCGACTGCCAGTCC  
CGCCAGCTCGGCAGGGCACGCCGTGGAAGTGCGGCCCGGGCTGTATCTGGGTGGGGCAGC  
GGCAGTGGCGGAACCGGGCCATCTGAGGGAGGCGGGCATCACCGCCGTGCTGACGGTGG  
ACTCTGAACCGGCTTTCCCGCTGGGGCTGGGTTTGAAGGTCTCCGGAGCCTCTTCGTGCC  
GGCGCTGGACAAACCCGAGACCGACCTGCTCAGCCACCTGGAATCGCTGCGTGGCCTTCAT  
CGGCCAGGCTCGCTCCGAAGGCCGCGCGGTGTTGGTGCACTGTTCATGCAGGAGTCAGTCG  
CAGTGTGTCTGTAGTGATGGCTTTTATAATGAAGACCGACAGCTTACCTTTGAAAAAGCC  
TACGACATCCTCCGACGGTCAAGCCAGAGGCTAAAGTGAATGAGGGGTTTGAATGGCAA  
CTGAAACTGTATGAGGCAATGGGATACGAAGTCGATACGTCCAGTGCCCTTTTACAAGCAG  
TACCGTTTACAAAAGGTGACTGAGAAGTATCCAGAATTTTGAATTTACCTCAGGAATCT  
TCGCTGTTGACCCAATACCATTTTACAGGGATTAAAGATGACATTCTTACAAATGCAG  
AAAAATGCAGGCGGTCTTTATTAGACATTCTAGTATTTTGGGTTCATAGTGAAGGAAGTGGG  
CCAATAGCCTTTGCTCACAAGAGAACGGCGCCATCTTCTGTACTTACCACAGGGAGTCAG  
GCTCAGTGCAGCTTCTTACTTCAATTGAGCCTGTGCAGTGGATGGAATCTACTCTGTTGGGCG  
TTATGGATGGACAGCTTCTTTGCCAAAATGCAGTGCCAAGTTGGGTTCTTTAACTGGTA  
TGGTGAAACAGTGTCTGTGTGGTTCGATGGATAACCCCTGCTTTTCAAATACACAAGAACAG  
AGTGGATGAAATGAAATGTTGCCGGCGCTGGGTTTACAGACAAAGAAGCTGTGAATTA  
GGACCCAGCTTGGGCTAGATCCTGTGAAAGGCACTTCCCGTGTGCATCATTCATGGCGGA  
TGTCAAACTTCTTTTGAATGTCCGAAGACAATTACTGACTGTAACAGCTGTGGGTTGGTA  
GCTTATTATGTGGCATATATACAGTTACGTTGCTTGGCAATTCAGATATTTAGTTATGTAA  
GTTGGATTGTTATTAAAACTTTTATAACCC

>SeqID\_17\_LOC51207\_h

GTGTTCCGCTCAGGCAGAGTCCTGCCCTGCACCCACTCCCCATTCCCGGCCCCAGGCCA  
TGCCCCAGGATGGACTACTGCAGAAGCAGGACCTCCGGAGGCCCAAGATCCATGGGGCA  
GTCCAGGCATCTCCCTACCAGCCGCCACATTGGCTTCGCTGCAGCGCTTGTGTGGGTCC  
GTCAGGCTGCCACACTGAACCATATCGATGAGGTCTGGCCAGCCTCTTCTGGGAGATG  
CGTACGCAGCCCGGGACAAGAGCAAGCTGATCCAGTGGGAATCACCCAGTTGTGAATG  
CCGCTGCAGGCAAGTTCAGGTGGACACAGGTGCCAAATTTCTACCGTGAATGTCCCTGG  
AGTACTATGGCATTGAGGCGGACGACAACCCCTTCTTCGACCTCAGTGTCTACTTCTGCC  
TGTTGCTCGATACATCCGAGCTGCCCTCAGTGTTCCCAAGGCCGCTGCTGGTACACTGT  
GCCATGGGGGTAAAGCCGCTCTGCCACACTTGTCTGGCCTTCTCATGATCTATGAGAACA  
TGACGCTGGTAGAGGCCATCCAGACGGTGCAGGCCCAACCGCAATATCTGCCCTAACTCAG  
GCTTCCCTCCGGCAGCTCCAGGTTCTGGACAACCGACTGGGGCGGGAGACGGGGCGGTTCT  
GATCTGGCAGGCAGCCAGGATCCCTGACCTTGGCCCAACCCACCAGCCTGGCCCTGGG  
AACAGCAGGCTCTGCTGTTTCTAGTGACCTGAGATGTAAACAGCAAGTGGGGGCTGAGG  
CAGAGGCAGGGATAGCTGGGTGGTGACCTCTTAGCGGGTGGATTCCCTGACCCAATTCA  
GAGATTCTTTATGCAAAAGTGAGTTCAGTCCATCTCTATAATAAAATATTATCGTCAT

>SeqID\_19\_AA813123\_h

GTGAGAGGAGACAGAAAGAGGGTGGTGGCCGATAGCTGGTCTCTTTCTCCAACACCTAG  
CCTGAGACTTGGCGGCGCGGTGCTATCCTGAACTAGCTTGGTAAGTGTTGTGTCCCGAAC  
CAGCGTAGAGAGACCTCGGACCAGCGCCTTGATGACAGCATCCGCGTCTCTTTTCATC  
ATCTCAGGGTGTCCAGCAGCCCTCCATCTACAGCTTCTCCCAAATAACCAGAAGCTTGT  
CTCAGCAATGGTGTGGCCGCCAACGACAACTCCTTCTGTCCAGCAATCGCATCACCGCC  
ATTGTCAATGCCTCGGTGGAAGTGGTCAACGTAATCTTCGAGGGCATTCACTACATAAAG  
GTGCCTGTTACCGATGCTCGTGACTCGCGTCTCTACGACTTTTTTGACCCATTGCTGATCT  
TATCCACACCATCGATATGAGGCAGGGCCGTACGCTGCTGCACTGCATGGCTGGAGTGAG  
CCGTTCCGCTCACTGTGCCTTGGTACCTCATGAAATACCACTCCATGTGCTGCTGGAC  
GCCCATACATGGACCAAGTCGCGCCGCCCATCATCCGGCCCAACAACGGCTTTTGGGAA  
CAGCTCATCAATTACGAATTCAAGCTGTTTAAATAACAACACCGTGCGCATGATCAACTCG  
CGGTAGGTAAACATCCCTGACATCTATGAGAAGGACCTACGTATGATGATATCAATGTAAG  
CCATCCCGGCCAGCCCTGACATCTGCCATCGATCTTGCAACCAAGACTGAACCTTTGAACAC

TGACATTTTGTAGTAAAGAAAACCGGATGGTGCCTTGTAAAGGGCAAGAAAAAGGGA  
GGGGGTTGGAGTTTGAACGTAGTAAGCCTTACCTTAATAGAATTAAATTCATGAAACAT

>SeqID\_21\_AA915932\_h

AATTACTTAGCGGCGACTGAGCCTATCGAGCAGTTTTCCATGGACACAGCCTAGCAGAAA  
GACGCAGCCTTCGTGCTTCGTGACTGCTGACCACTGACCCACCGCCTTGATGACAGCACC  
CTCGTGTGCCTTCCCAGTTCAAGTCCGGCAGCCCTCAGTCAGCGGCCTCTCGCAGATAACC  
AAAAGCCTGTATATCAGCAATGGTGTGGCCGCCAACAACAAGCTCATGCTGTCTAGCAAC  
CAGATCACCATGGTCATCAATGTCTCAGTGGAGGTAGTGAACACCTTGTATGAGGATATC  
CAGTACATGCAGGTACCTGTGGCTGACTCCCTAACTCACGTCTCTGTGACTTCTTTGACC  
CTATTGCTGACCATATCCACAGCGTGGAGATGAAGCAGGGCCGTAATTTGCTGCACTGTGC  
TGCTGGTGTGAGCCGCTCAGCTGCCCTGTGCCTCGCCTACCTCATGAAGTACCACGCCATG  
TCCCTGTGTGAGCGCCACAGTGGACCAAGTCAATGCCGGCCCATCATCCGACCCAAACAGC  
GGCTTTTGGGAGCAGCTCATCCACTATGAGTTCCAAATTGTTTGGCAAGAACAAGTGTGCACA  
TGGTCAGTTCCCCAGTGGGAATGATCCCTGACATCTATGAGAAGGAAGTCCGTTTGTATGAT  
TCCACTGTGAGCCATCCCACGAGCCCTGCATTGGAGTCAGAGGTACAGATCTATTGTTGA  
TCTTACACCAAGATCCAAACTTGAACATTCTACTTTTGTGATACAG

>SeqID\_23\_YVH1\_h

TGGGCGCGGCCATGTTGGAGGCTCCGGGCCCCGAGTGATGGCTGCGAGCTCAGCAACCCCA  
GCGCCAGCAGAGTCAGCTGTCCGGGCAGATGCTGGAAGTGCAGCCAGGATTGTATTTTCG  
GTGGGGCCGCGGCCGTCGCGGAGCCAGATCACTGAGGGAAGCGGGCATCACGGCCGTG  
CTAACAGTGGACTCGGAGGAGCCAGCTTCAAGGCGGGGCTGGGGTCGAGGATCTATGG  
CGCCTCTTCGTGCCAGCGCTGGACAAACCCGAGACGGACCTACTCAGCCATCTGGACCGG  
TGCGTGGCCTTCATCGGTGAGGCCCGCGCTGAGGGCCGTGCGGTGTTGGTGCATGTCTATG  
CAGGAGTCAGTCGAAGTGTGGCCATAATACTGCTTTTCTCATGAAGACTGACCAACTTCC  
CTTTGAAAAAGCCTATGAAAAGCTCCAGATTCTCAAACAGAGGCTAAGATGAATGAGGG  
GTTTGAGTGGCAACTGAAATTATACCAGGCAATGGGATATGAAGTGGATACCTCTAGTGC  
AATTATAAGCAATATCGTTTACAAAAGTTACAGAGAAGTATCCAGAATTGCAGAATT  
ACCTCAAGAACTCTTTGCTGTTGACCCAACTACCGTTTACAAGGATTGAAAGATGAGGTT  
CTCTACAAGTGTAGAAAGTGCAGGCGATCATTATTTCAAGTTCTAGTATTCTGGATCACC  
GTGAAGGAAGTGGACCTATAGCCTTTGCCACAAAGAGAATGACACCATCTTCCATGCTTA  
CCACAGGGAGGCAAGCTCAATGTACATCTTATTTTCAATGAACCTGTACAGTGGATGGAAT  
CTGCTTTGTTGGGAGTGTGGATGGACGCTTCTTTGCCAAAATGCAGTGCCAAAGTTGGG  
TTCCTTCAACTGGTATGGTGAACAGTGCTCTTGTGGTAGGTGGATAACACCTGCTTTTCAA  
ATACATAAGAATAGAGTGGATGAAATGAAAATATTGCCTGTTTTGGGATCACAAACAGGA  
AAAATATGAACATGATATTTATAGCTTGGGAAGAACTTGCAGATGATATGTGCTGCCTT  
TGCTTCTTATCATTCATGGCAGATTGTTAGTGCTTTCAACATTTCAATTTGAAATGGGAGAA  
GATAAAATCACTTGATGTAACCTGGAAACTATGCTTTACATGGCAATCAAAGCCTTTTGAT  
CATGTACATTTTATTTGATATTAATACTTTTATAACCAGAAA

>SeqID\_25\_SGP033\_h

GGGCGCCTGAGCCCTATATAGATCCTCAGGGCCCAGAAGCAGACTCTTCGGCGGGCGC  
CATGGGACCGTCAGAAAGTGGGCGCCGCGGGGCCGCTCGCCCGTACCGCCACCGTTGGT  
GCGCGTCGCGCCCTCACTCTTCTCGGGAGCGCGAGCCGCGGGCGCGGAGGAGCAGCT  
GGCGCGCGGGGAGTCACTCTGTGCGTCAACGTCTCCCGCCAGCAGCCCGGCCGCGCGC  
GCCCGCGTGGCAGAGCTGCGCGTGGCCGTGTTGACGACCCGGCTGAGGACCTGCTGGC  
GCACCTGGAGCCACGTGCGCCGCCATGGAGGCCGCGGTGCGCGCCGGCGGCGCCTGCCT  
AGTCTACTGCAAGAACGCGCCGAGCCAGCTCGGCGCCGTCTGCACCGCGTACCTCATGCG  
GCACCGCGGCTCAGCCTGGCGAAGGCCTTCCAGATGGTGAAGAGCGCTCGCCCGGTAGC  
AGAACCGAACC CGGGCTTCTGGTCTCAGCTCCAGAAGTATGAGGAGGCCCTCCAGGCCCA  
GTCCTGCCTGCAGGGAGAGCCCCAGCCTTAGGGTTGGGCCCTGAGGCTTGAAGCTTGAA  
GGCCTGCTGCCTGGAGGAAGGATGTCCCTGCACTGATACAGAAGGCTGGTCTTACCCCTTC  
TTCCTCACTGTCAATATCGAGTTTTCC

>SeqID\_27\_AI031656\_h

ACCTGGGCAATAAGGGACTAGCAGTTCAGCCGTTTTCTATGCCTGCTGGATTGTGTTGTAT  
TTGTTCCAGCCACTGCTCATGTAATGTACTCCCTTAACCAGGAAATTAAAGCATTCTCCC  
GGAATAATCTCAGGAAGCAATGCACCAGGGTGACAACGCTAACTGGAAAGAAAATTATA  
GAAACATGGAAAGATGCCAGAATTCATGTTGTGGAAGAAGTAGAGCCGAGCAGTGGGGG

TGGTTGTGGTTATGTGCAGGACCTTAGCTCGGACCTGCAAGTTGGCGTTATTAAGCCATGG  
TTGCTCCTAGGGTCACAAGATGCTGCTCATGATTTGGATACACTGAAAAAGAATAAGGTG  
ACTCATATTCTTAATGTTGCATATGGAGTTGAAAATGCTTTCCTCAGTGACTTTACATATA  
AGAGCATTTCTATATTGGATCTGCCTGAAACCAACATCCTGCTTATTTTCCAGAATGTTTT  
GAATTTATTGAAGAAGCAAAAAAGAAAGATGGAGTGGTTCTTGTTCATTGTAATGCAGGC  
GTTTCCAGGGGCTGCTGCAATTGTAATAGGTTTCTGATGAATTCTGAACAAACCTCATTTA  
CCAGTGCTTTTTCTTTGGTGAAAAATGCAAGACCTTCCATATGTCCAAATTCTGGCTTCAT  
GGAGCAGCTTCGTACATATCAAGAGGGCAAAGAAAGCAATAAGTGTGACAGAATACAGG  
AGAACAGTTCATGAGTTGCATTGTAGCAGACAATGGACAACCTGTAGTTTCTGAATTGACTT  
CTATAGCCATCTTTCCCTTTTGGAGAGTAGACTAGCAAAAATCCCTTTTTCTCTTGCTT  
TTTTATGCATAAAATGGAGGTCAATCTGATTGTCTGACCTACTGTATAAAG

>SeqID\_29\_MKP-5\_h

CACTATAGGGCGAATTGGGCCCTTAGATGCATGCTCGAGCGGCCCGCCAGTGTGATGGATA  
TCTGCAGAATTCCGCCTTACGATTTAGGTGACACTATAGAAGGTACGCCTGCAGGTACCG  
GTCCGGAATTCCCGGGTCGACCCACGCGTCCGCAATGAAGCCGAGTGAATGGGGGCTGAA  
TGTGCGAGTCCATAGCTGAAGAGGAGCGCCAGATGGTGGAGGAATACACTTATTTATGAA  
ACTGTCTTGAGTTCTTCTTGAATTGCCAGTTTTCAGCCTCCTCATGCCTCCGTCCTCTTAG  
ACGACAGGGTAGTAGTGGCACTATCTAGGCCCGTCCGACCTCAGGATCTCAACCTTTGTTT  
AGACTCTAGTTACCTTGGCTCTGCCAACCAGGCACTAACAGCCACCCTCCTGTCTATCGCC  
ACCACCGTTGTGTCCCTCAAGGCTGCGAATCTGACGTATATGCCCTCATCCAGCGGCTCTG  
CCCCTCGCTGAATTGTGGATGCAGCAGTGCCAGCTGCTGCACTGTGGCAACCTACGACA  
AGGACAAACAGGCCCAACCCAAAGCCATTGCCGTGGCACCACCACCTGCCATCGGAA  
CCTCTACCACCTGCCCTGCTAACAGATGGTCAACAATAATGAGAATACAGGCTCTCTAA  
GTCCATCAAGTGGGGTGGCGAGCCCTGTGTGTCAGGGACCCCAAGCAGCTAGCCAGCATCA  
AAATAATCTACCCCAATGACTTGGCAAGAAAGATGACCAATGCAGCAAGAGTCACTGTC  
CGAGTCAGGGCCCTGTCTATCTGACTGCAGGCCCTTCTATGGAGTACAACAAGAGTCA  
TCCAAGGAGCTGTCCACATTAACCTGTGCCGATAAGATCAGCCGGCGGAGACTGCAGCAG  
GCAAGATCACTGTCTAGACTTGATTTCTGTAGGGAAGGCAAGGACTCTTTCAAGAGGA  
TCTTTTCCAAAGAAATTATAGTTTATGATGAGAATACCAATGAGCCAAGCCGAGTGATGC  
CCTCCAGCCACTTCACATAGTCTCGAGTCCCTGAAGAGAGAAGGCAAGAACCTCTGG  
TGTTGAAAGGTGGACTTAGTAGTTTAAAGCAGAACCATGAAAACCTCTGTGACAACTCCCT  
CCAGCTCCAAGAGTGCCTGGGAGGTGGGGGGCGGCGCATCCGGGGCCTCGAGCTTGCTACC  
TCAGCCCATCCCCACCCCTGACATCGAGAACGCTGAGCTCACCCCATCTTGCCCTTC  
CTGTTCTTGGCAATGAGCAGGATGTCAAGGACCTGGACACCATGCAGCGGCTGAACATC  
GGCTACGTCATCAACGTCACCACTCATCTTCCCTCTACCACTATGAGAAAGGCCTGTTC  
ACTACAAGCGGCTGCCAAGCACTGACAGCAACAAGCAGAACCTGCGGCAGTACTTTGAAG  
AGGCTTTTGAAGTTCATTGAGGAAGCTCACCAAGTGTGGGAAGGGGCTTCTCATCCACTGCCA  
GGCTGGGGTGTCCCGCTCCGCCACCATCGTTCATCGCTTACTTGATGAAGCACTCGGATG  
ACCATGACTGATGCTTATAAATTTGTCAAAGGCAACGACCAATTATCTCCCCAACCTTA  
ACTTCATGGGGCAGTTGCTAGAGTTCGAGGAAGACCTAAACAACGGTGTGACACCGAGAA  
TCCTTACACCAAAGCTGATGGGCGTGGAGACGGTTGTGTGACAATGGTCTGGATGGAAAG  
GATTGCTGCTCTCCATTAGGAGACAATGAGGAAGGAGGATGGATTCTGGTTTTTTTTCTTT  
CTTTTTTTTTGTAGTTGGGAGTAAAGTTTGTGAATGGAAACAAACTTGGTTAAACACTTT  
ATTTTAAACAAGTGAAGAAGACTATCTTTTGTAGCCATTGAGATTACCTTCCACAAAC  
TGGCCAAATTAAGGAGGTTNAAGAAGTAATTTTTTTAAGCCCAACCATTAATAAATTAAT  
ACAACTTGGTTTCTNCCCCTTTTCTTTAAAGCTANTTTGTAAAAGTTTATGAG

>SeqID\_31\_NP\_060746\_h

ATGCTGCTGCTGGTGGCACAGCGGGACCGAGCCTCCCGCATCTTCCCCACCTCTACCTGG  
GCTCAGAGTGGAAACGCAGCAAACCTGGAGGAGCTGCAGAGGAACAGGGTCAACCCACATC  
TTGAACATGGCCCGGAGATTGACAACTTCTACCTGAGCGCTTCACTACCACAATGTGC  
GCCTCTGGGATGAGGAGTCCGGCCAGCTGCTGCCGCACTGGAAGGAGACGCACCGCTTCA  
TTGAGGCTGCAAGAGCACAGGGCACCCACGTGCTGGTCCACTGCAAGATGGGCGTCAGCC  
GCTCAGCGGCCACAGTGTGGCCTATGCCATGAAGCAGTACGAATGCAGCCTGGAGCAGG  
CCCTGCGCCACGTGCAGGAGCTCCGGCCCATCGCCGCCCCAACCTGGCTTCTGCGCCA  
GCTGCAGATCTACCAGGGCATCCTGACGGCCAGCCGCCAGAGCCATGTCTGGGAGCAGAA  
AGTGGGTGGGGTCTCCCCAGAGGAGCACCCAGCCCTGAAGTCTCTACACCATTCCCAAC  
TCTTCCGCCAGAACCTGAGGGTGGTGGGAGGAGAAAGTTGTAGGCATGGAAGAGAGCC  
AGGCAGCCCCGAAAGAGAGCCTGGGCCACGGCCACGTATAAACCTCCGAGGGGTCATG

AGGTCCATCAGTCTTCTGGAGCCCTCCTTGGAGCTGGAGAGCACCTCAGAGACCAAGTGAC  
ATGCCAGAGGTCTTCTCTCCACGAGTCTTACATGAAGAGCCTCTGCAGCCCTTCCCAC  
AGCTTGCAAGGACCAAGGGAGGCCAGCAGGTGGACAGGGGGCCTCAGCCTGCCCTGAAG  
TCCCGCCAGTCAAGTGTACCTCCAGGGCAGTGGCGTGGTGGCCAAACCGACCCAGGCC  
TTCCAGGAGCAGGAGCAGGGGCAGGGGCAGGGGCAGGGGAGAGCCCTGCATTTCTCTAC  
GCCAGGTTCCGGAAGGTGGTGAGACAGGCCAGCGTGATGACAGTGGAGAGGAGGGGC  
AGGCCTGA

>SeqID\_33\_NP\_060232\_h

TTACGCCCGGAGGCGTCGGCGCTGCCACTGGCCCCGACGGGAACGGGGCGAAAAAGGCG  
GCGGCACCATGTTCTCCCTCAAGCCGCCCAAACCCACCTTCAGGTCTACTTCTGCCACC  
GCCCCAGACTGACGATAAGATCAATTCGGAACCGAAGATTAATAAACTGGAGCCAGTCCT  
TTTGCCAGGAGAAAATGTGCTAAATGAAGTCAATTTGTGAGAAAAATGCATTGCAACAGA  
CACAAGCCAGTACGATTTGTGGGGAAAGCTGATATGCAGTAACCTCAAATCTCCTTTATT  
ACAGATGACCCAATGCCATTACAGAAATTCATTACAGAAACCTTCTTCTGGTGAACACG  
ATGTCCCTTAACATGTATTGAACAAATGTACAGTAAACGACCACAAGAGGAAGCAGA  
AAGTCTAGGCCCAACAGAACTGAAATTTAATCCAACAGAGTTAATTATTATTGTAA  
AGATTTGAGAATGTGAGATTTCGCTTTGATGAATCAGGTCCCGAAAGTGCTAAAAAGGT  
ATGCCCTTGCAATAGCTCATTATTCCAGCCAACAGACCTCCAGTACTCTTTGCATTTGAA  
TATGTTGGGAAAAAATACCACAATTCAGCAAAACAAAATTAATGGAATTCCTCAGGAGAT  
GGAGGAGGAGGAGGAGGAGGAGGTAATGGAGCTGGTGGTGGCAGCAGCCAGAAAACTCC  
ACTCTTTGAAACTTACTCGGATTGGGACAGAGAAATCAAGAGGACAGGTGCTTCCGGGTG  
GAGAGTTTGTCTATTAACGAGGGTTACATGATATCCACTTGCCTTCCAGAATACATTGTA  
GTGCCAAGTTCCTTTAGCAGACCAAGATCTAAAGATCTTTCCATTCTTTTGTGGGAGAA  
GGATGCCACTCTGGTGTGGAGCCACTTAACGGCAGTGCTCTGTGCGAATGGCCCTCAT  
CAAAGACGTGCTGCAGCAGAGGAAGATTGACCAGAGGATTTGTAATGCAATAACTAAAA  
GTCACCCACAGAGAAGTGATGTTTACAAATCAGATTGGATAAGACCTTGCTTAATATTCA  
AGAAGTACAAGCAGCATTGTAAAACTGAAGCAGCTATGCGTTAATGAGCCTTTTGAAGA  
AACTGAAGAGAAAATGGTTATCTTCACTGGAAAAATACTCGATGGTTAGAATATGTAAGGC  
ATTCTTAAGCATTACAGAGAATCTGTATACATGCTAGAAAGCAAACATCTCTGTAGTC  
CTACAAGAGGAGGAAGGAAGAGACTTGAGCTGTGTGTAGCTTCTTGTTCAGTGATG  
CTGGATCCCTATTTTAGGACAATTACTGGATTTCAGAGTCTGATACAGAAGGAGTGGGTCA  
TGGCAGGATATCAGTTCTAGACAGATGCAACCATCTAAAGAGATCAGAGAAAGAGTCTC  
CTTTATTTTGTCTATTCTTGGATGCCACCTGGCAGCTGTAGAACAAATATCTGCAGCTTTT  
GAGTTCTCCGAAACCTACCTGGCAGTGTTGTATGACAGCACCCGGATCTCACTGTTTGGCA  
CCTTCTGTTCAACTCCCCTCACCAGCGAGTGAAGCAAAGCACGGAATTTGCTATAAGCA  
AAAACATCCAATTTGGGTGATGAGAAGGGCTTAAAAATCCCCTCTGTTGGGACTGGTCTCT  
CCAGTTTACAGCAAAGGATCGCACCTTTTCCATAACCCCTTCTACATTGGAAGAGCACA  
CCTTGATACAGAAATGGCTCCGTGAAGTCTTTAAACGGACAAAGAAAAGCTACAGCTCC  
ACACTAAGAGGAATGCCGTCTGCCTTAAAGAAATGGAATCATCAGTGACCAAGAATTACTT  
CCAAGGAGAAATTCATTGATATTAACCAAGCCAGATCCAGCTCAGCAAACCGACAGC  
CAGAACAGTGATACGGAGCAGTATTTAGAGAATGGTTTTCCAAACCCGCCAACCTGCAC  
GGTGTATTCTGCCACGTGTCTCTGGAACACACATAAACTGTGGAACTGTGCTACTTCC  
GCTGGGTTCCCGAGGCCAGATCAGCCTGGGTGGCTCCATCACAGCCTTTCACAAGCTCTC  
CCTCCTGGCTGATGAAGTCGACGTACTGAGCAGGATGCTGCGGCAACAGCGCAGTGGCCC  
CCTGGAGGCTGCTATGGGGAGCTGGGCCAGAGCAGGATGTACTTCAACGCCAGCGGCC  
TCACCACACCGACACCTCGGGGACACCGGAGTTTCTCTCCTCCTCATTTCATTTTCTCCTG  
TAGGGAATCTGTGCAGACGAAGCATTATAGGAACACCATTAAAGCAAATTTTAAAGTGGGG  
CCAAAATATGGTTGTCTACTGAGACATTAGCAAATGAAGACTAAAAATAGGGTGTCTTCTG  
AACATTTTGAGGGAAGCTGTCAACTTTTCTCTGAATTAACATTGCTAACCTAGGCGTT  
TGAATCTCTAATAACTTTATATGTAAGAATAATAGTTGGAATTTGCACTAATATTAAAAA  
CATGTTGAATCATGCTTCTTTCACACTTATTTTAAAGAGAGATGTAAATTTTGTCTGTCT  
CTTTCTGTCAATTACAGGTCTGGCTCTTGTAACCGTGATCAAACCTGTTCAATGTTGTCTGCTAC  
ATTTTGTCTCCATCCATTTTCTTACCACCTCCTGAAGGCTATCTGATAGTCAGTCACATT  
AGCAGCCCCAGGCAGCAGACAACAGGAAAGTTAGGAAATTTGTGTTTCGTGTCATTTTA  
GGAGCATCTGATAAAACCTCCAGCAGGTTTTAGGAAGTATTCATGTATTTTCTGGTTACT  
TTCTGTCTCTCTAATTGAACTACCTGATGAAGGTTCAAGTTCTGGGGCCAGAAATTTAT  
GATTTTAGATCACCTTCTTTGAAACCTTAGATCACTGTGTTTTGAAATCATGAGTTGCTTT  
TAACTTCATAGGGTCAACTTTAAATGATATGCACTGTTAATTTTAAAGCATTGTGCTGAG  
ATAATTAACCTTAGAAGTGCCTTTGACTTTAGGATACAAATATTACAGAAGAAAAATATAA

FIGURE 4 (6 OF 8)

TTTCACITTTTAAAAATTGGGGTGGGAAAA\*CCCATTCGATATTTGAAATAGGCTTTTCATA  
CTAAGCTTCATAGCCAGGAGTCCCCAGAGTCTTGTTCCTCTGAAAGCCACTGGGGAGTGG  
CCTCTGGGGTGTGATTCCACAGAGGTGTATGCTGTAGACAGGAGAGTGCCATCTATGCC  
AAAACCTCGCCCTCAAAAAACAAACAAGGCTTGCTGGGAGGCGTGTGGGCTTGGCCATCAG  
TATTTCCAGTGTGGTAAACTATTGCTGGCACTTCCCCCTGGAAATAACTAATGAGGTTACG  
AGTTGGGCACCTGCACAGATGTCCTTCTCTCATAGTTCCTAATGCTTAGGAATAGAGGAGA  
AATAAAAAAATGGATTCTCTCAAAACACTGCCATTTGAATAGCGACAGAAGTGCTCCCCC  
AGCCCCAACTTTGGACAGCAAAGTTGAGGAGAAATGAGCAGACACAGTTGTTTGCTTGAT  
CTGAATCTCTCTAAAGTAAAGTATTTCCAACTGTGTGACAAGAGCCTACCTACCACTGTA  
GCGGTCAAAGCTGAAGCTTCTTACAGCAGTGAAACGGGGCACCACCTCCCCACACTCCT  
CATTTCCCGCTTAAAAACATGGATACTTTCAAATTTGACTGTTTCTTAACTGCCATCCTAA  
GATATGGAAAAATTTTTATAGTAAAGTGTCTAGTTAGCTTATTTCTTTCTAAAAACAAGTG  
TTTTCAAGATAACTGTATTTTACCTTTATATGTAAGTGAATAGCTGTTTCTTTTGAATTATTT  
GCCTTTTAAAAATTTGATAATGTCTCTGGATATAACAGGACAGGAGTCTTAAAAAATATCT  
TAAGAAATTCACCTTTATGGGTAAACCAAGGTTTTTGCCAACTTGTTCCTAGAA

>SeqID\_37\_MTMR7\_h

TGGTTCCTGCATGGTTCCTCCAAAGCTGCAGGAGGCCTTTGAACCTTTTCGATTTGAAACAT  
GCTGGTGCACACTTCAGAGCTCCGCCAGAGAATCCCTCGACCACGGGAGAACCGAGTC  
TTTCGGGGATTTGCCCTCCAGACAAGAGAAATGAGCAAGCAGGGAGCTCATCAGCTGTT  
GTTAGTGTATTTTACGTGTGTGGTATGGCCCAATACAGTTCGTCTTCTTCCAGTGTGGCTCA  
GGGAAGCCGAAAGGTTGAAAATGTCCGCTTGGTAGATCGAGTGTCTCTAAAAAAGCAGC  
TCTAGGTACTTTGTATTTGACGGCTACCCATGTCTATTCGTGGAAAAATTCACCTGACGCA  
AGAAAAGAAACATGGATTCTTCACAGTCAGATTTCCACCATTGAGAAACAGGCAACAACC  
GCTACCGGATGCCCTCTGCTGATTCGCTGCAAGAAGTTCAGATAATACAGCTCATCATAC  
CTCAGGAAGAGATTGCCACGACGTGTACATCTCCCTGATACGCCTTGCAAGGCCAGTGA  
AATATGAGGAGTTTACTGCTTTTCATTCAACCCCATGCTGGATAAAGAAAGAGAGAGC  
AAGGCTGGGTGCTGATCGATCTTAGTGAAGAATACACGCGGATGGGCCTCCCTAATCATT  
ACTGGCAGCTCAGCGATGTGAATAGAGACTATAGAGTCTGTGACTCTTATCCTACTGAACT  
GTACGTTCCCAAATCGGCCACGGCACACATCATAGTGGGGAGTTCCAAATTCGGAGTAG  
ACGGCGATTTCTGTCTTCTTCTTACTATTATAAAGATAACCACGCTCCATCTGCCGGAGC  
AGCCAGCCCCCTGTCCGGCTTCAAGTCCCCGGTGCCTAGAGGACGAGCAGATGCTCCAGGCC  
ATTAGGAAAGCCAATCCAGGAAGTGACTTCGTTTATGTCTTGACACCCGGCCTAAACTTA  
ATGCAATGGCAAATCGTGCTGCAGGGAAGGCTATGAGAATGAAGACAATTATTCGAATA  
TCAAGTTTCAGTTTATCGGGATAGAGAACATCCATGTCTAGGGAACAGTCTGCAGAAAA  
TGCTGGAAGTGTGTGAACCTTAAATCTCCCTCCATGAGTGATTTCTGTGGGGTCTGGAGAA  
CTCTGGCTGGTTAAGGCACATTAAGCCATAATGGATGCAGGAATCTTCATTGCAAAGGC  
AGTGTGAGAGGAAGGGGCAAGTGTGCTTGTCACTGTTCTGATGGCTGGGACAGGACCGC  
TCAGGTGTGCTCGGTGGCAAGCCTGCTGCTGGACCCCTCACTACCGGACTCTGAAGGGCTC  
ATGGTATTAATTGAAAAGGACTGGATTTCTTTGGTCAAAAGTTTAAATCACCGATATGGCA  
ATCTAGATGGTGACCCAAAAGAAATCTCTCCAGTTATTGACCAGTTCATTGAGTGTGTTTG  
GCAGTTAATGGAACAATTTCCCTGTGCCTTTGAGTTCAATGAGAGGTTTTTGATTACATT  
CAACATCACATTTATTCTGCCAGTTTGGAACTTCCTATGTAACAGCCAAAAGGAGAGA  
CGAGAACTCAAGATTCAAGAAAGAACATACTCATTATGGGCTCACCTGTGGAAGAATCGG  
GCCGACTACCTGAATCCTCTGTTTAGAGCTGATCACAGCCAGACTCAGGGAACCTTTTCATC  
TCCCTACAACACCATGTAACCTCATGTACAAGTTTTGGAGTGGAAATGTATAACCGCTTTGA  
AAAGGGGATGCAGCCCCGACAGTCAGTTACAGATTACCTAATGGCAGTGAAGGAAGAAA  
CTCAGCAGCTAGAGGAAGAACTAGAGGCCCTGGAAGAAGTAAGACATACTTGTCTTGTTA  
ATCTATTTCTGTGCTTATTTCTTGAAGAGCAAGGCTTCATGAAAGCCTCTTAAGATTAGT  
GCAATACAAATATTTACATTTTATGTCAGTAGAACACCTGAAACACAACCCCATCCAGGT  
CAGGACTCAGATAACCGAAGCCAAACCTAGGCATTTAATGCATTTACACATCTGGTTTG  
TGGCTCAGTTCACAAAATTTTGGTTGTTCTTATTTAGGTCAACTGTTTGGTTACTGTT  
ATTTGAGAGTAGTAAATTAATTTTTTCTGGAAAGTGTGAAAAAGTAACAGAATAAAAT  
CAAATATATGTAAGTAAACTCTTTCCCACTTTTAGAAATATGTTAACTATGATATTA  
ATCTATAATTGGTAAAGGGGAGACTGACTGTAAAGTCTTATCTTTTGGCTTCTATGAACCT  
CCCTGTACTTGGACAGGGAGGTTATAATAAATGCTTTTCTAATTGGAA

>SeqID\_39\_AA493915\_h

ATGAACGAAAGTCTGATCCGAATGCCCTGGCGGGAGTCATCATTGAGAGAGCCCCAGT  
GACAGTCCACAGACAAACGAATTTAAAGGAGCAACCGAGGAGGCACCTGCGAAAGAAAG

CCCACACACAAGTGAATTTAAAGGAGCAGCCTGGTGTACCTATCAGTAAAAGTATGTT  
AGAACGACTTTCCAAGTTTGAAGTTGAAGATGCTGAAAATGTTGCTTCATATGACAGCAA  
GATTAAGAAAATTTGTGCAATTCAATTGTATCATCCTTTGCATTTGGACTATTTGGAGTTTCC  
TGGTCTTACTGGATGTCACTCTCGTCCCTTGCCGACCTAATTTTCACTGACAGCAAACCTTAT  
ATTCTTTCGGAGTATCGTTCTATTTCTCTAGCTATTGCCTTATTTTTTCTCATGGATGTTCTT  
CTTCGAGTATTTGTAGAAGGGCCCGTCTATACCATTTGGGCTGCCCCCTTCAGACCTCCGGG  
CAGGAAAAGAAGAGACTGTACTGGTCAGAGAAAAGACATCAGCAGGAGAGCCAGAGATTTC  
CTCTCTCTCCATCATCACCATCACCATTATTCTCATCACCATCACCGTCACCGTCATCAA  
ATATTTTAATTTAACTAAAAATATTAACCTTGAAATCAGTAAGATGGTGGTATTCTCCAAG  
GAAGTAAATGAGTGGATGACTCAAGATCCTGAAAACATCATAGTGATTCACTGTAAAGGA  
GGCAAAATCATCATCACCATCATGGACTTCAAAGAAGTTGTACAACTCAATATTGCAAA  
GTTGTCAAGTTCTCTCAAGTTAATCTATAAAATCAATGTAGTTCCAATAAAAAATCCTAAATG  
TGAAAGGAAGAACCGGAACCTATGGTTTGTGCGCTCCTTATTGCCTCCGAAATATTTTAAAC  
TGCAGAGGAAAAGCCTGTATTATTTGGAGAAAGGCGAACAGATAAAAACCCACAGCAATA  
AATTTCAAGGAGTAGAAAACCTCTTGTGAGAATAGATATGTTGGATATTTTGCACAAGTGA  
AACATCTCTACAACGGGAATATCCCTCCAAGACGGATACTCTTTATAAAAAGATTCAATTAT  
TTATTCGACTCGTGGTGTGGAACAGGTGATGTATGTGATCTACAATTCCAAATAGTAATG  
GAGAAAAAGGTTGTCTTTTCCAGTACTTCATTAGGAAAATTGTTTCGCTG

>SeqID\_41\_BAA91172\_h

GTGGCCCGGGAGGCGCCGAGGCCAGGTAGGTGCGATGGGCGTGCAGCCCCCAACTTCTC  
CTGGGTGCTTCCGGGCGGGCTGGCGGGACTGGCGCTGCCGCGGCTCCCCGCCCACTACCA  
GTTCTGTGGACCTGGGCGTGCGGCACCTGGTGTCCCTGACGGAGCGCGGGCCCCCTCAC  
AGCGACAGCTGCCCCGGCCTCACCTGCACCGCTGCGCATCCCCGACTTCTGCCCCGCGG  
CCCCCGACCAGATCGACCGCTTCGTGCAGATCGTGGACGAGGCCAACGCACGGGGAGAG  
GCTGTGGGAGTGCATGTGCTCTGGGCTTTGGCCGCACTGGCACCATGCTGGCCTGTTACC  
TGGTGAAGGAGCGGGGCTTGGCTGCAGGAGATGCCATTGCTGAAATCCGACGACTACGAC  
CCGGCCCCATCGAGACCTATGAGCAGGAGAAAGCAGTCTTCCAGTTCTACCAGCGAACGA  
AATAAGGGGCCTTAGTACCCTTCTACCAGGCCCTCACTCCCTTCCCCATGTTGTCGATGG  
GGCCAGAGATGAAGGGAAGTGGACTAAAGTATTAAACCCTCTAGCTCCCATTTGGCTGAAG  
ACACTGAAGTAGCCACCCCTGCAGGCAGGTCTGATTGAAGGGGAGGCTTGTAAGTCTT  
TGTTGAATAAATGAGTTTTACGAACCAGGGA

>SeqID\_2\_AA023073\_m  
MGVQPPNFSWVLPGRLAGLALPRLPAHYQFLLDQGVRLVSLTERGPPHSDSCPGTLHRMRI  
PDFCPSPEQIDQFVKIVDEANARGEAVGVHICALGFGRTGTMLACYLVKERALAAGDAIAEIR  
RLRPGSIETYEQEKAVFQFYQRTK

>SeqID\_4\_AA030322\_m  
MGTSEAAPPPFARVAPALFIGNARAAGATELLVRAGITLCVNVSRQQPGPRAPGVAELRVPVF  
DDPAEDLLTHLEPTCAAMEAAVRDGGSCLVYCKNGRSRAAVCTAYLMRHRGHSLDRAFM  
VKSARPVAEPNLGFWAQLQKYEQTLOAQAILPREPIDPE

>SeqID\_6\_AA103595\_m  
MSGMSQILPGLYIGNFKDARDAEQLSRNKVTHILSVHDTARPMLEGVKYLCIPAADTPSQNL  
TRHFKESEIKFIHECRLQGESCLVHCLAGVSRSVTLVIAIYIMTVDFGWEDALHTVRAGRSCANP  
NLGFQRQLQEFKEHEVHQYRQWLREEYGENPLRDAEEAKNILAAPGILKYWAFRLRL

>SeqID\_8\_AA144705\_m  
MDSLQKQELRRPKIHGAVQVSPYQPPTLASLQRLWVRRATLTHINEVWPNLFLGDAYAAR  
DKGRLIQLGITHVNVNAAAGKFQVDTGAKFYRGTPLEYGIEADDNPFDLVHFLPVARYIRD  
ALNIPRSRVLVHCAMGVSRSAITVLAFLMIFENMTLVDAIQTQVAHRDIPNSGFLRQLQVLD  
NRLRRETGRL

>SeqID\_10\_AA274457\_m  
MHSLNQEIKAFSRDNLKQCTRVTLTGKKLIETWEDATVHVVEPSSGGGCGYVQDLTLD  
LQVGVIKPVWLLGSQDAAHDLLELRKHKVTHILNVAYGVENAFLESEYTKTISILDVPETNLS  
YFPECFEIEQAKLKDGVLVHCNAGVSRAAIVIGFLMSSEATFTTALSIVKEARPSICPNPG  
FMEQLRTYQVGKESNGGDKVPAEDTTGGL

>SeqID\_12\_AA374753\_h  
MGNGMNKILPGLYIGNFKDARDAEQLSKNKVTHILSVHDSARPMLEGVKYLCIPAADSPSQNL  
TRHFKESEIKFIHECRLRGESCLVHCLAGVSRSVTLVIAIYIMTVDFGWEDALHTVRAGRSCANP  
NVGFQRQLQEFKEHEVHQYRQWLKEEYGESPLQDAEEAKNILAAPGILKFVAFRLRL

>SeqID\_14\_AA396428\_m  
EPLVLKGGSSSFQNHGNLCDNSLQLECREVGGGASAASSMLPQSVPTTPDIENAELTPILPF  
LFLGNEQDAQDLDTMQRLNIGYVINVTTHLPLHYHEKGLFNYKRLPATDSNKQNLRYFEEA  
FEFIEEAHQCGKGLLIHCQAGVSRSAITVIAIYLMKHTRMTMTDAYKFVKGKRPIISPNNFMG  
QLLEFEDDLNNGVTPRILTPKLMGMETVV

>SeqID\_16\_AA422661\_m  
MLEAQGTNHGCERQAPTASPASSAGHAVEVRPGLYLGGAAVAEPGHLREAGITAVLTVDS  
PAFPAGAGFEGRLSLFVPALDKPETDLLSHLDRCVAFIGQARSEGRAVLVHCHAGVSRSAVV  
MAFIMKTDQLTFEKAYDILRTVKPEAKVNEGFEWQLKLYEAMGYEVDTSYAFYKQYRLQKV  
TEKYPELWNLPELFAVDPTTISQGLKDDILYKCRKCRRLFRHSSILGHSEGSGPIAFHKRTA  
PSSVLTGSGAQCTSYFIEPVQWMESTLLGVMDGQLLCPKCSAKLGSFNWYGEQCSCGRWITP  
AFQIHKNRVDEMMLPALGSQTKKL

>SeqID\_18\_LOC51207\_h  
MDSLQKQDLRRPKIHGAVQASPYQPPTLASLQRLWVRQAATLNHIDEVWPSLFLGDAYAAR  
DKSKLIQLGITHVNVNAAAGKFQVDTGAKFYRGMSLEYGIEADDNPFDLVSYFLPVARYIRA  
ALSVPQGRVLVHCAMGVSRSAITVLAFLMIYENMTLVEAIQTQVAHRNICPNSGFLRQLQVL  
DNRLGRETGRF

>SeqID\_20\_AA813123\_h  
MTASASSFSSSQGVQPSIYSFSQITRSLFSLNGVAANDKLLSSNRITAIVNASVEVVNVFFEGI  
QYKVPVTDARDSRLYDFFDPIADLIHTIDMRQGRITLLHCMAGVSRASLCLAYLMKYHSM  
LDAHTWTKSRRPPIRPNNGFWEQLINYEFLFNNTVRMINSPVGNIPDIYEKDLRMISM

>SeqID\_22\_AA915932\_h



MTAPSCAFFVQFRQPSVSGLSQITKSLYISNGVAAANNKMLLSSNQITMVINVSVEVVNTLYEDI  
QYMQVPVADSPNSRLCDFDPIADHHSVEMKQGRITLLHCAAGVSRSAALCLAYLMKYHAM  
SLDAHTWTKSCRPIIRPNSGFWEQLIHYEFQLFGKNTVHVMVSSPVGMIPIYEKEVRLMIPL

>SeqID\_24\_YVH1\_h

MLEAPGPSDGCESNPSASRVSCAGQMLEVQPGLYFGGAAAVAEPDHLREAGITAVLTVDSE  
EPSFKAGPGVEDLWRLFVPALDKPETDLLSHLDRCVAFIGQARAEGRAVLVHCHAGVSRVAI  
ITAFMLKTDQLPFEKAYEKLQILKPEAKMNEGFEWQLKLYQAMGYEVDTSIAIKYRLQKV  
TEKYPELQNLQELFAVDPTTVSQGLKDEVLYKCRKCRSLFRSSILDHREGSGPIAFHKRM  
TPSSMLTTGRQAQCTSYFIEPVQWMESALLGVMDGQLLCPKCSAKLGSFNWYGEQCSCGRWI  
TPAFQIHKNRVDEMKILPVLGSQTGKI

>SeqID\_26\_SGP033\_h

MGPSEAGRRGAASPVPPPLVRVAPSLFLGSARAAGAEQLARAGVTLCVNVSRQQPGPRAPG  
VAELRVPVFDPAEDLLAHLEPTCAAMEAAVRAGGACL VYCKNGRSQLGAVCTAYLMRHR  
GLSLAKAFQMVKSARPVAEPNPGFWSQLQKYEEALQAQSQLQGEPPALGLGPEA

>SeqID\_28\_AI031656\_h

MYSLNQEIKAFSRNNLRKQCTRVTTLTGKKIETWKDARIHVVEEVEPSSGGGCGYVQDLSSD  
LQVGVIKPWLLGSDAAHDLDTLKKNKVTHILNVAYGVENAFLSDFTYKSSILDLPETNILS  
YFPECFEFIEEAKRKDGVLVHCNAGVSRAAIVIGFLMNSEQTSFTSAFSLVKNARPSICPNNG  
FMEQLRTYQEGKESNKCDRIQENS

>SeqID\_30\_MKP-5\_h

MPPSPLDLDRVVVALSRPVRPQDLNLCDDSSYLGSANPGSNHPPVIATTVSLKAANLTYMPS  
SSGSARSLNCGCSSASCCTVATYDKDNQAQTQAIAAGTTTTAIGTSTTCPANQMVNNNTGTS  
LSPSSGVGSPVSGTPKQLASIKIYPNDLAKKMTKCSKSHLSQGPVIDCRPFMEYNKSHIQGA  
VHINCADKISRRLQOGKITVLDLISCREGKDSFKRIFSKEIIVYDENTNEPSRVMPSPQLHIVLE  
SLKREGKEPLVLKGGSSFKQNHENLCDNSLQLOQECREVGGGASGASSLLPQPIPTTPDIENAE  
LTPILPFLFLGNEQDVRDLDTMQLNLIGYVINVTTHLPLHYEKGFLFNYKRLPSTDSNKNQNLRO  
YFEEAFEFIEEAHQCGKGLLIHCQAGVSRSAITVIA YLMKHTRMTMTDAYKFVKGKRPIISPNL  
NFMGQLLFEEDLNNGVTPRILTPKLMGVETVV

>SeqID\_32\_NP\_060746\_h

MLLLVAQRDRASRIFPHLYLGSEWNAANLEELQRNRVTHILNMAREIDNFYPERFTYHNVR  
WDEESAQLLPHWKETHRFIEAARAQGTHVLVHCKMGVSRSAATVLA YAMKQYECSEALR  
HVQELRPIARPNPGFLRQLQIYQGILTASRQSHVWEQKVGVSPEEHPAPEVSTPFPPLPEPEG  
GGEEKVVGMEESQAAPKEEPPRPRIINLRGVMRSISLLEPSLELESTSETSDMPEVFSSHESSE  
EPLQPFQQLARTKGGQVDRGPQALKSRQSVVTLQGSAAVVRNRTQAFQEQQGQGGQGGEP  
CISSTPRFRKVVVRQASVHDSGEEGEA

>SeqID\_34\_NP\_060232\_h

MFSLKPPKPTFRSYFLPPPQTDDKINSEPKIKKLEPVLLPGEIVVNEVNFVRKCIATDTSQYDLW  
GKLICSNFKISFITDDPMPLQKFHYRNLLGEHDVPLTCIEQIVTVNDHKRKQKVLGPNQKLKF  
NPTELIIYCKDFRIVRFRFDESGPESAKKVCLAIAHYSQPTDLQLLFAFEYVGKKYHNSANKING  
IPSGDGGGGGGGGGAGGGSSQKTPLFETYSDWDREIKRTGASGWRVCSINEGYMISTCLPEY  
IVVPSSLADQDLKIFSHSFVGRMPLWCWSHNSGALVRMALIKDVLQQRKIDQRICNAITKSH  
PQRSDVYKSDLDKTLNIEQVQAAVFKLKQLCVNEPFEETEEKWLSSLENTRWLEYVRAFLK  
HSAELVYMLESKHLVSVLQEEGRDLSCCVASLVQVMDPYFRITTFQSLIQKEWVMAGYQ  
FLDRCNHLKRSEKESPLFLFLDATWQLEQYPAAFEFSETYLA VLYDSTRISLFGTFLFNSPHQ  
RVKQSTEFASKNIQLGDEKGLKFPVWDWSLQFTAKDRTL FHNPFYIGKSTPCIQNGSVKSKF  
RTKKSYSSTLRGMPALKNGUSDQELLPRRNSLILKPKPDPAQQTDSQNSDTEQYFREWFSKP  
ANLHGVILPRVSGTHIKLWKLCYFRWVPEAQISLGGSSITAFHKL SLLADEVDVLSRMLRQQR  
GPLEACYGELGQSRMYFNASGPHHTDTSGTPEFLSSSFPFSPVGNLCRRSILGTPLSKFLSGAKI  
WLSTETLANED

>SeqID\_38\_MTMR7\_h

MVPPKLQEAPEFDLKHAGAHFRAPPRESLDHREN RVFRGFAPPDKRNEQAGSSSAVVSFVYV  
CGMAQYSSSSSSVAQGSRKVENVRLVDRVSPKKAALGTYLTATHVIFVENSPPDARKETWILH  
SQISTIEKQATTATGCPLLIRCKNFQIQLIIPQERDCHDVYISLIRLARPVKYEELYCFSFNPMIL

FIGURE 5 (2 OF 3)

KEEREQGWVLIDLSEYYTRMGLPNHYWQLSDMNRDYRVCDSTPTELYVPKSATAHIIVGSSKF  
RSRRRFPVLSYYYKDNHASICRSSQPLSGFSARCLEDEQMLQAIRKANPGSDFVYVVDTRPKL  
NAMANRA IGKGYENEDNYSNIKFQFIGIENIHVMRNSLQKMLEVCELKSPSMSDFLWGLENS  
GWLRIKAIMDAGIFIAKAVSEEGASVLVHCSDGWDRTAQVCSVASLLLDPHYRTLKGFMVLI  
EKDWISFGHKFNHRYGNLDGDPKEISPVIDQFIECVWQLMEQFPCAFEFNERFLIHQHHIYSCQ  
FGNFLCNSQKERRELKIQERTYSLWAHLWKNRADYLNPLFRADHSQTQGTFLPTTPCNFMY  
KFWSGMYNRFEKGMQPRQSVTDYLMVKEETQQLLEELEALEEVRHTCFVNLFSVLIS

>SeqID\_40\_AA493915\_h

MNESPDPNALAGVIERAPSDSPQTNEFKGATEEAPAKESPHTSEFKGAALVSPISKSMLERLSK  
FEVEDAENVASYDSKIKKIVHSIVSSFAFGLFGVFLVLLDVTLLADLIFTDSKLYPSEYRSISL  
AIALFFLMDVLLRVFVEGPVYTIGLPPSDLRAGKEETVLVRERHQESQRFLLLSITTTIITITTV  
TVIKYFNLTKNIKLEISKMVVFSKEVNEWMTQDPENIIVHCKGGKIITIMDFKEVCTTQYCKV  
VSSLKLIYKFNVVPIKILNVKGRGTGMVCALLIASEIFLTAESLYYFGERRTDKTHSNKFQGV  
TPCQNRVYGYFAQVKHLYNGNIPPRILFIKRFIYSTRGVGTGDVCDLQFQIVMEKKVVSSTS  
LGNC SL

>SeqID\_42\_BAA91172\_hMGVQPPNFSWVLPGRLAGLALPRLPAHYQFLDLGVRLVSLTERG  
PPHSDSCPGLTLHRLRIPDFCPPAPDQIDRFVQIVDEANARGEAVGVHICALGFGRTGTMLACYL  
VKERGLAAGDAIAEIRRLRPGPIETYEQEKAVFQFYQRTK

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**